

THE MATRICES OF LINEAR LEAST-SQUARES IMAGE ANALYSIS

By LOUIS GUTTMAN

The Israel Institute of Applied Social Research, Hebrew University, Jerusalem¹

By the 'image' of a variable is meant the multiple regression of that variable on all the other variables in the data to be analysed, and its 'anti-image' the error of estimate. It is then shown that an expression for R , the ordinary matrix of covariances or correlations, can be obtained in terms of the covariance matrices of the images and anti-images respectively. Various theorems and identities are deduced which lead to new structural theories for R and for data in the social sciences generally. In particular fresh light is thrown on the problems of parsimony and rank.

I. IMAGES AND ANTI-IMAGES

Consider a set of n quantitative variables $x_1, x_2, \dots, x_n$ for a given population of individuals. The purpose of image analysis is to help analyse the structure of the statistical interrelations of the variables for the population. When all regressions are linear, they are all functions of the observed covariance matrix R . When regressions are not linear, then R does not give complete information about the interrelations, and some more general image technique is required than the present one. The most general in principle is Israel Alpha, which was described for a special qualitative case in [10], but which holds for more general cases, both qualitative and quantitative. The present paper is devoted to the linear quantitative case called 'Israel Beta' [9]. This has proved especially helpful both in studying basic problems of common-factor theory [5, 12, 13, 14] and in the radex approach to factor analysis [2, 6, 7, 11].

If all observed variances are unity, R will be the observed correlation matrix. In image analysis, as it turns out, it makes no essential difference in what units the observed variances are expressed; so we shall let these variances be arbitrary, and not necessarily equal to unity.

By the image x_j^* of x_j in the given set of data is meant the prediction of x_j from its multiple regression on the remaining $n-1$ variables. If w_{jk} denotes the multiple regression coefficient of x_k for predicting x_j , then

$$x_j^* = \sum_{k=1}^n w_{jk} x_k \quad (j=1, 2, \dots, n), \quad (1)$$

it being understood that

$$w_{jj} = 0 \quad (j=1, 2, \dots, n), \quad (2)$$

since a variable is not used to predict itself. The anti-image of x_j , denoted by e_j , is defined to be the errors of estimate

$$e_j = x_j - x_j^* \quad (j=1, 2, \dots, n). \quad (3)$$

¹ This research was facilitated in part by a grant from the Lucius N. Littauer Foundation.

II. THE FUNDAMENTAL IDENTITY

A fundamental identity was established in [5] for studying the structure of R . Let s_j denote the standard error of estimate of x_j from the remaining $n-1$ variables, or

$$S^2_j = \text{var}(e_j) \quad (j=1, 2, \dots, n); \quad (4)$$

and let S denote the diagonal matrix whose j th main diagonal element is s_j ,

$$S = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \dots & \\ & & & s_n \end{bmatrix}. \quad (5)$$

Let G and Γ denote the Gramian covariance matrices of the images and the anti-images respectively. Then

$$R = G - \Gamma + 2S^2. \quad (6)$$

It has also been shown in [12, 13] how this identity (6) is helpful for studying the implications of a common-factor analysis in the sense of Spearman, Thurstone, Burt, and others. Such a factor analysis seeks a diagonal matrix U , a Gramian matrix L , and a weight matrix A such that

$$R = ALA' + U^2, \quad (7)$$

where, as usual, the prime indicates a transposed matrix. In many cases, especially when the number of common factors is small compared with n , $S^2 \rightarrow U^2$ as $n \rightarrow \infty$ [13, 14].

Some important paradoxes arise in considering the limits as $n \rightarrow \infty$, and contrasting these limits with what will hold for finite n . The purpose of the present paper is to facilitate exploring these paradoxes by establishing some universal properties or tautologies for G , Γ , and S^2 for finite n . For example, it is of interest to ask whether S^2 cannot possibly serve as a U^2 for finite n as well as in the limit. The answer turns out to be 'no'. Let R_1 be defined as the difference

$$R_1 = R - S^2. \quad (8)$$

R_1 is the same as R except in the main diagonal. We shall prove the following theorem.

Theorem 1. If R is a nonsingular, but not a diagonal, matrix, then R_1 cannot be Gramian for finite n , where R_1 is defined by (8). But this does not prevent the limit of R_1 from being Gramian as $n \rightarrow \infty$.

Thus, not all of the $n s_j^2$ can be uniquenesses simultaneously, for $R - U^2_j$ must be Gramian in (7). However, in certain special cases, some of the s_j^2 determined by finite n may serve as uniquenesses [13]. Further points of interest concern the rank of G and related matrices. Its rank is the dimensionality of the image space, or of what the observations have in common in the image sense. One of our surprising results is that G and R_1 always have the same rank

when R is nonsingular. Indeed, we shall prove three other matrices of importance have the same rank as G . These additional matrices will be non-Gramian, but will turn out to have the same *signature* as R_1 . Furthermore, we shall show that G can have any rank from 2 to n inclusive, but not rank 1. This does not prevent the rank of the *limit* of G from being 1 as $n \rightarrow \infty$. Finally, the concept of 'parsimony' in factor analysis will be reviewed in the light of the identities of image analysis.

III. THE COVARIANCE MATRICES AND SOME FURTHER IDENTITIES

For each subscript j , we have three kinds of variables in our analysis: x_j , x_j^* , and e_j . We are interested in the covariances within and between each of these kinds. If z_j is one of these variables, it will be convenient to regard z_j as a vector in a Euclidean vector space, and to define the inner product between two vectors z_j and z_k as the *covariance* between z_j and z_k , denoting it by $z_j z_k'$. Since a covariance does not change if the variables are reversed in order, $z_j z_k' \equiv z_k z_j'$. In particular, $z_j z_j'$ is the variance of z_j .

This inner product convention enables us to deal with continuous variables as well as with the discrete ones implied by the ordinary matrix algebra used so frequently for this purpose in factor analysis [12]. Indeed, our theorems hold for the general, abstract, finite-dimensional, Euclidean vector space, of which the present statistical terminology is but one interpretation. With this convention, the typical elements of R , G , and Γ are $x_j x_k'$, $x_j^* x_k^*$, and $e_j e_k'$ respectively:

$$R = [x_j x_k'], \quad G = [x_j^* x_k^*], \quad \Gamma = [e_j e_k']. \quad (9)$$

We can define further the covariance matrices *between* the different kinds of vectors:

$$R_1 = [x_j x_k^*'], \quad S^2 = [x_j e_k'], \quad H = [x_j^* e_k']. \quad (10)$$

According to (10) the typical element of S^2 is the covariance between observed x_j and the anti-image of x_k . Does this clash with definition (5) for S ? As is well known in least-squares theory [cf. 17], a necessary and sufficient condition for the anti-images actually to have minimum variance is that S^2 be a diagonal matrix as implied by (5), or

$$x_j e_k' = e_j x_k' = 0 \quad (j \neq k). \quad (11)$$

Thus, e_j is an 'outer' or 'vector' product of the $n-1$ predictors, being a vector orthogonal to their $n-1$ subspace. On postmultiplying both members of (1) all through by e_j' , and using (11) and (2), we at once confirm the well-known property that the image and anti-image of a given variable are orthogonal to each other:

$$x_j^* e_j' = 0 \quad (j=1, 2, \dots, n); \quad (12)$$

for x_j^* is a vector lying in the subspace of the predictors, being the projection of x_j on that subspace. Postmultiplying (3) through by e_j' and using (12) confirms the further well-known property that

$$e_j e_j' = x_j e_j' \quad (j=1, 2, \dots, n). \quad (13)$$

Properties (11) and (13) show that the definition of S^2 in (10) coincides with definition (5) for S .

Similarly, we shall show that the definition of R_1 in (10) coincides with that in (8). This does not seem to have been noted before. The typical element of R_1 in (10) is the covariance of observed x_j with the image of x_k . Postmultiply both members of (3) by x'_k and use the definitions of S^2 , R , and R_1 respectively in (10) and (9); we then see that $S^2 = R - R_1$, from which (8) follows.

Although its definition in (10) does not indicate this at first blush, R_1 must always be a symmetric matrix according to (8), i.e. $x_j x'_k = x_k x'_j$. Image analysis shows how the concept occurs naturally of modifying the main diagonal of R —much as in common-factor theory—by means of formula (8) for the natural matrix R_1 defined in (10).

IV. IDENTITIES INVOLVING H

The typical element of H in (10) is the covariance of the image of x_j with the anti-image of x_k . H is related to the previously defined matrices by the identity

$$H = R_1 - G. \quad (14)$$

To prove this, postmultiply (3) all through by x'^*_k , and recall the definitions of the matrices of (14). G is certainly symmetric by definition, indeed Gramian; and the symmetry of R_1 has already been noted. Hence (14) shows that H must also be symmetric. Thus we can write

$$R'_1 = R_1, \quad H' = H. \quad (15)$$

Another identity of interest is

$$\Gamma = S^2 - H. \quad (16)$$

To establish this, postmultiply (3) through by e'_k and recall the definitions of Γ , S^2 , and H in (9) and (10). Using (8) and (14) in (16) and then transposing terms provide an alternative proof of identity (6). A particular way of rewriting (6), which will be of more direct relevance to us later, is

$$R - S^2 = G + (S^2 - \Gamma), \quad (17)$$

which is simply a transposition of the terms of (14), or

$$R_1 = G + H. \quad (18)$$

V. THE CASE OF NONSINGULAR R ; THE MATRIX W

Given definitions (9) and (10) and the normal eqns. (11), all of formulae (5), (6), (8), (12)–(18), are identities, with no assumptions whatsoever. Many facts can be deduced about rank and signature for our covariance matrices, but only in terms of inequalities unless assumptions are now introduced. The case of greatest empirical interest is where R is nonsingular; and for this important equalities hold. From now on, therefore, we shall restrict ourselves for brevity to the nonsingular case.

One matrix is essentially defined above, but not yet discussed, namely, the multiple regression weights, W ,

$$W = [w_{jk}]. \quad (19)$$

This is a square matrix of order n , with zeros down the main diagonal. Postmultiplying (1) by x'_i shows that

$$R_1 = WR. \quad (20)$$

Postmultiplying (20) by R^{-1} and recalling (8) we have

$$W = I - S^2 R^{-1}, \quad (21)$$

where I is the unit matrix of order n . This last identity supplies one way of verifying three important and well-known propositions:

- (a) W is in general asymmetric;
- (b) s_j^2 is the reciprocal of the j th diagonal element of R^{-1} ;
- (c) A necessary and sufficient condition that R be nonsingular is that S^2 be nonsingular.

Proposition (a) follows by observing that the second term on the right of (21) is asymmetric in general. Proposition (b) can be seen from the main diagonals in (21)—which show that $w_{jj} = 1 - s_j^2 r^{jj}$ where r^{jj} is the j th main diagonal element of R^{-1} —and recalling (2) and the fact that no main diagonal element of a nonsingular Gramian matrix can vanish [3, p. 56]. Thus, if we define H_0 as

$$H_0 = S^{-2} - R^{-1}, \quad (22)$$

then the main diagonal elements of H_0 are all zero, and we can write

$$\text{diag}(W) = \text{diag}(H_0) = 0. \quad (23)$$

The necessity part of proposition (c) is implied by proposition (b); the sufficiency follows by noting from (20) and (8) that

$$S^2 = (I - W)R \quad (24)$$

so that the nonsingularity of the left member implies the nonsingularity of both $I - W$ and R . Thus, S^{-2} exists wherever R^{-1} does, and conversely. The further consequence that $I - W$ must be nonsingular when R is leads to *Theorem 2. If R is nonsingular, then no latent root of W can equal unity; $I - W$ must be nonsingular.*

Further identities involving W include

$$G = W'RW, \quad (25)$$

which follows from postmultiplying each member of (1) by its own transpose, revising subscripts; and

$$H = WS^2. \quad (26)$$

To verify (26) postmultiply (1) by e'_i , and recall (10). Expanding the right-hand member of (26) by (21) shows that

$$H = S^2 - S^2 R^{-1} S^2, \quad (27)$$

which upon comparison with (22) yields

$$H = S^2 H_0 S^2. \quad (28)$$

Remembering (23), and using (26) and (28), we can verify that H is a hollow matrix, with zero main diagonal elements in accordance with (12), i.e.,

$$\text{diag}(H) = 0. \quad (29)$$

VI. THE BASIC THEOREM ON RANK

Having assembled our *dramatis personae* and their interrelations, we can now state and prove our basic theorem on rank.

Theorem 3. If R is nonsingular, then all of the following matrices must have the same rank: G , H , H_0 , R_1 , W .

The kingpin in the proof is W . Since S^2 is nonsingular by Proposition (c), (26) shows H has the rank of W , and therefore so does H_0 in (28). By a theorem on the transformation of a nonsingular Gramian matrix [3, p. 75], (25) shows that G has the rank of W . Identity (20) shows R_1 has the rank of W when R is nonsingular. Hence the five matrices of the theorem all have the same rank.

A curious consequence is as follows. From (16) and (27),

$$\Gamma = S^2 R^{-1} S^2, \quad (30)$$

i.e. Γ is nonsingular whenever R is. However, from (16), $\Gamma - S^2$ has the rank of H , which is the rank of G and of R_1 according to Theorem 3. In other words, the covariance matrix of the *anti-images* is nonsingular when R is, but modifying its main diagonal by subtracting out S^2 reduces its rank to the rank of the *image* space, or to whatever rank R is reduced to by subtracting S^2 . It is interesting to note that in (18), the rank of the sum is the rank of each term in the sum.

VII. THE BASIC THEOREM ON SIGNATURE

While $R - S^2$ and $\Gamma - S^2$ have the same rank as G , they differ from G in an important respect: they are not Gramian (except in the trivial case where $G = R - S^2 = 0$, i.e. where all the x_j have zero correlation with each other). If we exclude G from the quintet of matrices in Theorem 3, we can establish the following:

Theorem 4. If R is nonsingular, then the matrices H , H_0 , R_1 , and W all have the same respective numbers of positive and negative latent roots; and the symmetric matrices H , H_0 , and R_1 all have the same signature.

The proof depends upon Sylvester's 'law of inertia' [8] and two lemmas. The first is well known [3].

Lemma 1. If A and B are square matrices of the same order, then AB and BA have the same latent roots (or spectrum).

It follows that HS^{-2} has the same latent roots as $S^{-1}HS^{-1}$. Since $W = HS^{-2}$ from (26), W has the same spectrum as $S^{-1}HS^{-1}$. But by Sylvester's law the signature of $S^{-1}HS^{-1}$ is the same as that of H . Hence, the frequencies of

positive, negative, and zero latent roots of W must be respectively equal to those of H . (This incidentally shows that all latent roots of W must be real, even though W is asymmetric.) By Sylvester's law it also follows from (28) that H and H_0 have the same signatures, and hence the same distributions of signs for the latent roots.

The case of R_1 remains to be considered. For this, we introduce

Lemma 2. If G_1 and G_2 are nonsingular Gramian matrices of the same order, then the rank and signature of $G_1 - G_2$ are the same respectively as of $G_2^{-1} - G_1^{-1}$. This lemma, so far as is known, is new. For proof, let F be a nonsingular matrix satisfying

$$G_1 = FF', \quad (31)$$

and let G_0 be defined by

$$G_0 = F^{-1}G_2(F^{-1})'. \quad (32)$$

Then, by expansion, the following can be easily verified as identities:

$$F^{-1}(G_1 - G_2)(F^{-1})' = I - G_0, \quad (33)$$

and

$$F'(G_2^{-1} - G_1^{-1})F = G_0^{-1} - I. \quad (34)$$

Now, the signatures of $I - G_0$ and $G_0^{-1} - I$ must be equal, for if λ is a latent root of G_0 , then $1 - \lambda$ and $\lambda^{-1} - 1$ are latent roots respectively of $I - G_0$ and $G_0^{-1} - I$, and $1 - \lambda$ and $\lambda^{-1} - 1$ are clearly of the same sign for all positive λ . Sylvester's law and (33) show that the rank and signature of $G_1 - G_2$ are respectively those of $I - G_0$. Similarly, the rank and signature of $G_2^{-1} - G_1^{-1}$ are respectively equal to those of $G_0^{-1} - I$. We have just seen that $I - G_0$ and $G_0^{-1} - I$ have the same rank and signature, and consequently Lemma 2 follows. According to Lemma 2, $R - S^2$ and $S^{-2} - R^{-1}$ have the same rank and signature, i.e. R_1 and H_0 satisfy Theorem 4. This completes proof of Theorem 4.

Theorem 1 above is a corollary of Theorem 4. Since the trace of H_0 (or of W , or of H) is zero by virtue of (23), the sum of its latent roots must also be zero. Hence if H_0 has at least one positive root, it must also have at least one negative root. Therefore the same is true of R_1 ; i.e. Theorem 1 follows. A further corollary is stated in Theorem 5.

Theorem 5. If R is nonsingular, then the rank of G (or of H , H_0 , R_1 , and W) cannot be unity.

H_0 , for example, cannot be of rank one; for, if it has a latent root of one sign, it must also have a root of the opposite sign, or at least two nonvanishing latent roots. Hence, if it does not vanish, H_0 must have rank two at least; therefore so must G and the other members of the quintet of Theorem 3.

The theorems above are for finite n . It has been shown in [5, 12] that they need not hold for the limits of the matrices involved as $n \rightarrow \infty$. Further consequences of this paradox, especially for the concept of rank in factor analysis, will be explored in a later article.

VIII. A CONSTRUCTION OF R

Structural theories for R can be developed in terms of hypotheses about the interrelations of images and anti-images. We shall illustrate this briefly here by considering a construction based on H . Let H be any symmetric matrix of order n with zeros in the main diagonal. Clearly, if $H \neq 0$, H can have any rank from 2 to n . Let S be any nonsingular diagonal matrix such that $S^2 - H$ is Gramian and nonsingular. A sufficient, but not a necessary, condition is that each diagonal element of S^2 exceed the largest latent root of H . Let $\Gamma = S^2 - H$, and let $R = S^2 \Gamma^{-1} S^2$. Then R is a nonsingular covariance matrix for which H , Γ , and S^2 satisfy the definitions given in (9) and (10) and all the identities of this paper.

If we desire to express the structure of a matrix of *correlations*, and not merely covariances, we can take D to be the diagonal matrix whose j th diagonal element is the positive square root of the corresponding element of R constructed in the preceding paragraph; then $D^{-1}RD^{-1}$ will be the corresponding correlation matrix; and in the preceding paragraph, we can accordingly replace H , S^2 , and Γ by $D^{-1}HD^{-1}$, $D^{-2}S^2$, and $D^{-1}\Gamma D^{-1}$ respectively. This illustrates the effects of changing observed standard deviations in an image analysis. The rank and signature of H , and hence the rest of our quintet, remain unchanged.

IX. POSSIBLE STRUCTURAL THEORIES

Focusing on H can lead to new kinds of structural theories for social science data. For example, if h_{jk} is the typical element of H , one possibility is that h_{jk} is of the simple form

$$h_{jk} = (a_j - a_k)^2 \quad (j, k = 1, 2, \dots, n), \quad (35)$$

where a_j is some coefficient associated with observed variable x_j . Evidently (35) defines a symmetric matrix with zeros in the main diagonal, and so provides a possible H . Expanding the right-hand member shows that

$$h_{jk} = a_j^2 + a_k^2 - 2a_j a_k \quad (j, k = 1, 2, \dots, n). \quad (36)$$

From the right-hand member of (36), we see that H is the sum of three matrices, each of rank one, so that H as defined by (35) is in general of rank 3.

More generally, if

$$h_{jk} = (a_j - a_k)^{2p} \quad (j, k = 1, 2, \dots, n), \quad (37)$$

then we have a possible H defined for each positive integer p , and of rank $2p + 1$ in general.

An *order* theory results from considering an H defined by

$$h_{jk} = a_j - a_k \quad (j \leq k; j, k = 1, 2, \dots, n). \quad (38)$$

In contrast to (35), we see that in (38) the order in which a_j and a_k are subtracted from each other makes a difference; (38) implies an ordering of the x_j . And it is easy to show that if neighbouring a_j are all distinct in (38), then H is *non-*

singular. One way is to subtract each row of H in (38) from its successor, and each column in the result from its own successor; we then obtain for the determinant of H the equation

$$|h_{jk}| = 2^{n-2} (a_1 - a_n) (a_2 - a_1) (a_3 - a_2) \dots (a_n - a_{n-1}). \quad (39)$$

Clearly, a necessary and sufficient condition for $|h_{jk}|$ not to vanish is that the a_{j+1} be distinct from a_j , setting $a_{n+1} = a_1$.

A more general order model is given by the condition

$$h_{jk} = (a_j - a_k)^{2p-1} \quad (j \leq k; j, k = 1, 2, \dots, n), \quad (40)$$

where p is a positive integer. Still more general formulae for an H can be expressed as weighted sums of H 's of the forms (37) and (40) with varying values of p . Other possible formulae could be considered.

X. PARSIMONY VERSUS RANK

It is not our intention here to explore the possible psychological or sociological meaning of models such as (37) and (40) above. They may or may not be wholly inappropriate for any known kind of data. Our purpose in presenting them is twofold. First, they are among the simplest for illustrating possible laws of formation for H . Secondly, they cast further light on the important problem of *parsimony* in factor analysis.

In certain schools of factor analysis there has been some danger of equating the concept of 'parsimony' with that of 'small rank' for R [1, 4]. Thus many researches have been aimed at finding a U^2 for (7) that will make $R - U^2$ of relatively small rank compared with n , in the belief that this will yield the most parsimonious description of R . The models for R , based on H , which we have given above, show that a parsimonious description need have nothing to do with rank. In each of (35) and (38), we have a single common-factor theory for R ; only one parameter a_j is needed for each observed variable. These models are in a sense just as parsimonious as Spearman's original model. Indeed, for Spearman's case, it follows from Lemma 2 that $U^{-2} - R^{-1}$ is of rank one whenever $R - U^2$ is; and hence an H for Spearman's R must have the form

$$h_{jk} = a_j a_k \quad (j \neq k). \quad (41)$$

According to (41), since $h_{jj} \equiv 0$, H is in general nonsingular for Spearman's case, although all diagonal-free submatrices are of rank one. In general diagonal-free matrices of (35) are of rank three; those of (38) are of rank two if on one side of the main diagonal, being in general nonsingular if they contain elements from both sides of the main diagonal.

This illustrates the fact that models can be equally parsimonious in the number of parameters they require, and yet vary widely in the ranks they specify for R with modified main diagonal. This has been pointed out before in the special case of the simplex (in the radex approach to factor analysis)

where for example r_{jk} is of the form a_j/a_k ($j \leq k$); i.e. only a single parameter for each observed variable is required [4, 11]. The present analysis emphasizes once again that rank is not the only possible point of departure for factor analysis.

The identities of image analysis afford a basis for examining more closely the implications of rank in the Spearman-Thurstone type of common-factor-theory, and the adequacy of computing routines based on the rank concept. This important topic, however, requires a separate exposition.

REFERENCES

- [1] FERGUSON, G. A. (1954). The concept of parsimony in factor analysis. *Psychometrika*, XIX, 281-289.
- [2] GABRIEL, K. R. (1954). The simplex structure of the progressive matrices test. *Brit. J. Statist. Psychol.*, VII, 9-14.
- [3] GUTTMAN, L. (1942). Properties of Gramian matrices. Chapter two in *The prediction of quantitative varieties by factor analysis*. Doctoral dissertation, Univ. of Minnesota.
- [4] GUTTMAN, L. (1947). Review of Thurstone's *Multiple-factor analysis*. *J. Am. Statist. Assn.*, XLII, 651-656.
- [5] GUTTMAN, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, XVIII, 277-296.
- [6] GUTTMAN, L. (1953). Two new approaches to factor analysis. Office of Naval Research, Annual technical report on project Nonr-731(00), Washington, D.C.
- [7] GUTTMAN, L. (1954). A new approach to factor analysis: the radex. In Paul F. Lazarsfeld (ed.), *Mathematical thinking in the social sciences*. The Free Press.
- [8] GUTTMAN, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, XIX, 149-161.
- [9] GUTTMAN, L. (1954-5). An outline of some new methodology for social research. *Public opinion quarterly*, XVIII, 395-404.
- [10] GUTTMAN, L. (1954). The Israel alpha technique for scale analysis. In Riley, et al., *Sociological studies in scale analysis*, Rutgers Univ. Press.
- [11] GUTTMAN, L. (1955). A generalized simplex for factor analysis. *Psychometrika*, XX, 173-192.
- [12] GUTTMAN, L. (1955). The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. *Brit. J. Statist. Psychol.*, VIII, 65-81.
- [13] GUTTMAN, L. (1956). 'Best possible' systematic estimates of communalities. *Psychometrika* (in press).
- [14] GUTTMAN, L. Une solution au problème des communautés. *Bull. Centre. d'Etudes et Recherches Psychotechniques*, VI, 123-128.
- [15] GUTTMAN, L. (1958). What lies ahead for factor analysis? *Educ. and Psychol. Measurement*, XVIII, 497-515.
- [16] GUTTMAN, L. (1958). To what extent can communalities reduce rank. *Psychometrika*, XXIII, 297-308.
- [17] JACKSON, D. (1930). The theory of approximation. *Am. math. soc. colloquium publications*, XI.

THE ANALYSIS OF PERSONAL PREFERENCES

By PATRICK SLATER

Institute of Psychiatry, University of London

A procedure is explained for the construction of preference tables which set out in a systematic scheme the results obtained by using the method of ranks, paired comparison, or some other type of multiple comparison; and the detailed properties of such tables are described. Such a table may be used to exhibit the dispersion of objects in a multi-dimensional space of individuals' preferences or the dispersion of individuals in a space of preferences for objects; and the two devices are shown to be complementary. Methods of analysing and studying tables of this kind are discussed and illustrated.

I. THE CONSTRUCTION OF A PREFERENCE TABLE

Data of the kind to be discussed in the following paper may be arrayed in a preference table such as that illustrated in Table I. The figures here given are taken from an experiment in which nine objects were shown to 34 individuals for ranking in order of preference and to a further 28 for comparison in pairs [10]. The columns *A*, *B*, . . . , *E* refer to five of the objects, and the rows 1, 2, . . . 8 to eight individuals who compared them in pairs. The zero entry in row 1 under *A* indicates that the first individual did not choose *A* in preference to any of the other four objects paired with it; and the next entry in the same row indicates that he preferred *B* in comparison with three of the other four, and so on. In general if there are *m* objects and *n* persons a preference table, *T*, can be drawn up with *n* rows and *m* columns. A typical individual is denoted

TABLE I. PREFERENCE TABLE

Individual	Object				
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
1	0	3	2	1	4
2	3	4	1	0	2
3	4	0	3	1	2
4	3	2	4	1	0
5	3	4	2	0	1
6	1	2	0	3	4
7	3	1	4	2	0
8	1	2	0	4	3
Total	18	18	16	12	16

I, a typical object *O*, and a typical entry in the table t_{io} : this designates the entry in the row for *I* and the column for *O*; and is defined as the number of other

objects in comparison with which I prefers O . The row of entries for I is denoted (t_i) and the column of entries for O (t_o) . A similar table may be constructed when using a different method of comparison. The objects might be presented in a single set of m for ranking or in suitable numbers of sets of intermediate size between 2 and m , subject to certain restrictions. From the entries in Table I alone it would be impossible to tell whether they have been obtained by paired comparison or by some other method of comparison. The objects may vary in more than one respect; and in choosing between them the observer need not be directed to consider any one respect in particular. He can give any weight he likes to any differences he notices.

II. THE INTERNAL CONSISTENCY OF A SINGLE SCHEDULE OF RESPONSES

Let us consider first the properties of a single schedule of responses, from which the entries in a single row of T are obtained. Except when the method of ranking has been used, such a schedule will yield evidence for the internal consistency of the responses or their lack of it; the best evidence is provided when the method used is paired comparison: for, when the comparisons are made in pairs, all the evidence obtainable when they are made in sets of more than two is secured, with the addition of further evidence obtainable only with pairs.

The complete schedule of responses given by I in comparing m objects can be exhibited in a preference matrix, P_i , where the o th row and column is allotted to the object O . If I prefers J to K when these two objects are presented for comparison $+1$ is entered in row J column K and -1 in row K column J . The entries in the leading diagonal are all zero. Figure 1 exhibits the preference matrix for a schedule relating to five objects: units and zeros have been omitted to make inspection easier. P_i is a skew-symmetric or anti-symmetric matrix

FIGURE 1. A PREFERENCE MATRIX

	A	B	C	D	E	$(t_i)'$
A		+	+	+	-	3
B	-		+	+	+	3
C	-	-		+	+	2
D	-	-	-		+	1
E	+	-	-	-		1

with the property $P_i + P_i' = (0)$ where (0) is an $m \times m$ matrix with zeros everywhere. (t_i) , transposed, is obtained by summing the positive entries in each row. All the responses in Fig. 1 except one, $A < E$ (i.e. E preferred to A) are consistent with an arrangement of the objects along a scale of preference in the order $A > B > C > D > E$; and no other order is consistent with so many of the responses. It may be called the 'nearest adjoining order'. But reversing $A > C$, $A < E$, and $C > E$, (cf. Fig. 2) leaves (t_i) unchanged, although the resulting

P_i cannot have less than two inconsistent responses; and any of the orders $A < B < C < D < E$, $A < B < D < E < C$, $A < B < E < C < D$, $B < C < A < D < E$ or $C < A < B < D < E$ might serve. Thus information about the consistency and order may easily be lost when a schedule of responses is reduced to (t_i) . As m increases, the scope for internal variation in the P_s increases.

FIGURE 2. A PREFERENCE MATRIX

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	$(t_i)'$
<i>A</i>		+	—	+	+	3
<i>B</i>	—		+	+	+	3
<i>C</i>	+	—		+	—	2
<i>D</i>	—	—	—		+	1
<i>E</i>	—	—	+	—		1

The null hypothesis will state that the observer noticed no differences between the objects and simply picked one at random every time a pair was presented. The hypothesis is tested by counting the number of responses inconsistent with the nearest adjoining order, or any of the nearest adjoining orders if there are alternatives. The number is denoted by i . When $m=5$, the null hypothesis is still tenable even if $i=0$; but it can be rejected when i and m are 0 and ≤ 6 , 1 and ≤ 7 , 2 and ≤ 8 , 3 and ≤ 9 respectively.

Thus certain schedules conform beyond reasonable doubt with some ordering of the objects along a scale of preference. Such a scale is characteristic of the individual in question, and is called an I -scale. But even when P_i conforms with an I -scale, doubt may persist about the correct order. If I chooses $A > B$, $B > C$ and then $C > A$, an inconsistency is involved and two hypotheses are tenable concerning it: either I cannot distinguish between A , B and C and chooses one at random each time a pair is presented, or some of his choices are made in accordance with one preference scale and some in accordance with another. The relation between i and m implies that, provided m is large enough, any number of inconsistencies may be left to explain in a P_i , even after the operation of one I -scale has been demonstrated. If all are consistent with some other ordering of the objects, can their number be sufficient to demonstrate the operation of a second I -scale? The hypothesis that only one I -scale is needed to characterize an individual's choices may seem niggardly; but the evidence of a single P_i is never sufficient to disprove it. The i responses remaining after a nearest adjoining order has been found must always be consistent with a second or residual I -scale. But no psychological meaning need attach to it. Nor is it possible to disprove the null hypothesis (viz., that the inconsistencies originate from small independent causes, such as fortuitous distractions or momentary lapses of attention) without supplementing the evidence of the P_i .

III. PROPERTIES OF THE I -SCALE

The foregoing arguments present a case for reducing any P_i to a single vector defining a dimension of preference characteristic of an individual, not for preferring (t_i) to any other vector. But there are, as we have seen, strong grounds for preferring the nearest adjoining order, provided it can be defined unambiguously; and there are also grounds for preferring a vector of weights obtained by the iterative procedure due to Wei ([16], reported by Kendall [6]). None of these vectors can be claimed as optimal except in relation to some specific purpose. One merit not to be despised is 'face-validity'. The scoring procedure which gives (t_i) is simple and reasonable, and can be applied to P_i obtained by any method of multiple comparison. Wei's iterative procedure uses (t_i) as a first approximation; and it is always questionable whether the advantages of further refinement are sufficient to offset the additional labour involved.

(t_i) is a partition of $m(m-1)/2$ into m terms. Their mean must be $(m-1)/2$, and this must be true for all the t_{i0} in T . When the objects are ranked, the sum, v_i , of the squares of the deviations of the terms in (t_i) from their mean must be $(m+1)m(m-1)/12 = \phi_{im}$, but when other methods are used, v_i may be smaller. A P_i can never generate a $v_i > \phi_{im}$ and can only generate a $v_i = \phi_{im}$ when it has no inconsistencies. Inconsistencies always reduce v_i to some extent. When the comparisons are made in sets of x , ϕ_{ix} , the theoretically expected value of v_i , will be equal to $m(m-1)(x+1)/12$. When the comparisons are made in pairs, v_i is related to Kendall's d [5], the number of circular triads involved in the P_i : in fact, $d = \frac{1}{2}(\phi_{im} - v_i)$. Hence the properties of d noted by Kendall may be used to infer corresponding properties of v_i . This provides a further reason for using (t_i) to define an I -scale¹.

The usual preliminary standardization is not required; a much simpler procedure will suffice. All the (t_i) have the same mean and commensurate variances. When ranking is employed, all the v_i equal ϕ_{im} , and thus, without any further standardization, an equal weight is assigned to the observed variation on each I -scale. When some other method is employed, the v_i may vary in magnitude; but their size varies with the internal consistency of the P_i from which they are derived; so if we retain them as they are instead of re-scaling, we assign a weight to each I -scale which corresponds with its reliability.

Accordingly, before proceeding to the analysis, all that is needed is to set the origin of each I -scale at its mean by subtracting $(m-1)/2$ from each element in T . Using U to denote the dispersion with its origin at its multivariate mean point, we may define its characteristic element as $u_{i0} = t_{i0} - (m-1)/2$. The row of entries for I in U , namely, (u_i) , is given by $P_i \cdot (\frac{1}{2}) = (u_i)'$, where $(\frac{1}{2})$ is a column vector with m elements each equal to $\frac{1}{2}$. The U obtained from Table I is shown in Table II.

Each column specifies the co-ordinates of an object in the I -space, the n -dimensional continuum of individual preference scales. The contours of the region

¹ Kendall's d is not connected with i by any precise relation. Hence, a test for internal consistency in a P_i may lead to different conclusions when d and i are both used.

occupied by the objects are described by the I -matrix, the matrix of sums of squares and products of the readings on the I -scales, which can be defined algebraically as UU' . Table III gives I -matrix obtained from Table II.

TABLE II. U-TABLE BASED ON TABLE I

Individual	A	B	C	D	E
1	-2	1	0	-1	2
2	1	2	-1	-2	0
3	2	-2	1	-1	0
4	1	0	2	-1	-2
5	1	2	0	-2	-1
6	-1	0	-2	1	2
7	1	-1	2	0	-2
8	-1	0	-2	2	1

TABLE III. I -MATRIX
defining the dispersion of the objects in the I -space

Individual	Individual							
	1	2	3	4	5	6	7	8
1	10	2	-5	-5	0	5	-7	2
2	2	10	-5	1	9	-1	-3	-3
3	-5	-5	10	5	0	-5	6	-6
4	-5	1	5	10	5	-10	9	-9
5	0	9	0	5	10	-5	1	-6
6	5	-1	-5	-10	-5	10	-9	9
7	-7	3	6	9	1	-9	10	-7
8	2	3	-6	-9	-6	9	-7	10

Principal component analysis may be applied directly. Here $m=5$; hence there cannot be more than 4 non-zero latent roots and vectors. The dimension of preference defined by the major axis may be described as the one where the differences between the objects are most conspicuous or most commonly recognized, and so on.

IV. PROPERTIES OF THE O -SCALES

The entries in the column of T relating to O , denoted by t_o , may range from $(m-1)$ to zero. An entry of $(m-1)$ can only be obtained from an individual who prefers O to all the other objects; $(m-2)$ from an individual who prefers O to all except one; and so on. Thus (t_o) arrays individuals according to their preference for O ; and we now find ourselves taking a different view of T . Each (t_o) is seen as a vector defining a dimension of preference for an object, i.e. an O -scale, and each t_{io} locates an individual on this O -scale. Thus T , which previously specified a dispersion of m points in an n -space, now specifies a dispersion of n points in an m -space. No change has occurred in T itself, only

in our view of it. Both views are equally valid. The two dispersions may be described as *cognate*, since both are derived from T . Though not the same, their properties must be related. A correct and sufficient specification of the relation should permit the properties of either to be inferred from the other.

Before describing a relation of this kind, I would like to clear up some misconceptions about properties of the (t_o) . The distribution of t_{io} on an O -scale may assume a great variety of forms; and it is useful to distinguish between the forms to be expected (a) when the observers cannot discriminate between the objects, (b) when they show discernment without agreement, and (c) when they exhibit agreement. Whether there is discernment or not, so long as there is *no agreement* the distribution must tend to be the same for every object, i.e. the same on every O -scale as it is in T . And the distribution in T can be inferred from the properties of (t_i) . The mean of t_{io} is $(m-1)/2$, and the sum of squares of deviations is v_i . So in T the mean will also be $(m-1)/2$, and the sum of squares $\Sigma v_i = V$ (say). Hence on any O -scale the mean must tend to remain at $(m-1)/2$, and the sum of squares at V/m .

Evidence of *discernment in the absence of agreement* is to be found in the magnitude of V . It will tend to approach $n\phi_{im}$ when each observer tends to be consistent, and $n\phi_{ix}$ when he chooses at random. The opportunity for evidence of discernment to appear is thus greater the wider the difference between ϕ_{ix} and ϕ_{im} , i.e. the smaller x is. For testing significance it would be an advantage to have information on the probability integral of the ratio $V/n\phi_{ix}$ for given m and n when $x=2$. No existing procedure seems altogether satisfactory.

The same argument can be used to obtain details about the frequency distribution of the t_{io} on the hypothesis of *no discernment*. In this case the *probability* distribution of the t_{io} derived from a (t_i) can be obtained from the expansion of the multinomial $\{1/x + \dots + 1/x\}^{(m-1)(x-1)}$ where $\Sigma 1/x = 1$. $(m-1)/(x-1)$ must be an integer if the m objects are to be shown in sets of x . The expected frequency distribution of the t_{io} in T is derived by multiplying each term in the expansion by nm . In general it is unimodal and symmetrical, and flattens out as x increases. It terminates abruptly with frequencies of $nm x^{-(m-1)(x-1)}$ at the limiting values $t_{io}=0$ and $m-1$. These frequencies are not small under the usual conditions, i.e. when n is large, even if m and x are relatively small. The distributions approach the normal at a point far beyond the region of practical interest, viz., when m increases indefinitely, and n and x are kept small. Within the region of practical interest they resemble the rectangular form found when $x=m$. This is also the form when there is *discernment without agreement*: for, if there is discernment, each observer's responses are internally consistent and the (t_i) generated by each P_i is a permutation of the natural numbers from 0 to $m-1$; and if there is no agreement, no one of these permutations is more likely than any other.

Thus certain specific forms are assumed by the distributions when there is no agreement. *And the remaining forms—an almost innumerable variety—must imply agreement of one kind or another.* This may come as a surprise to students

of psychology or statistics; but it will seem no novelty to students of economics or law.

The simplest form of agreement encountered in psychological experiments may be described as limited agreement on a single irreversible order. This form has monopolized attention since experiments based on multiple comparison began in the nineties. We may suppose that the observers generally agree that A should be placed above B , B above C , etc., and that the divergences of opinion are inconsistent with one another and all slight. Then O is generally assigned a particular place in the order, and, if it is not assigned that place exactly, it is more likely to be assigned one near it than one more remote. All the distributions of t_{io} will tend to be unimodal. For columns with modes near the general mean they will also tend to be symmetrical; as the modes move towards either end of the scale, the distributions will become more skewed. For the objects generally ranked first or last they are likely to be J -shaped. They resemble the family of binomial distributions generated by expanding $(p+q)^{m-1}$ when p is allowed to assume a succession of values between 0 and 1 and $q=1-p$; but they are noticeably flatter. But the cumulative frequency distribution, i.e. the distribution of t_{io} in T , will tend to be rectangular even when it is not constrained to be.

Without postulating more than a single order we may now relax the restriction of irreversibility progressively, and consider how the means, variances, and covariances on the O -scales will be affected. Suppose that the observers, when asked to rank the objects, tend to arrange them in an alphabetical order $A, \dots, M$, but when asked, "Which is the top?" some point to A and some to M . There is agreement on an order, but it is *reversible*, or as Burt would say, bipolar. We could mark off points for $A, \dots, M$ on a single scale and define the observed order in an individual case in terms of the distance between these points and a point p which may occupy two alternative positions, either at A or at M . Comparing the entries for individuals 4, 6, 7 and 8 in Table I, the reader may suspect that the data there provide some evidence for a prevalent bipolar order.

But why restrict p to two alternative positions? Let us allow it to wander anywhere along the scale between A and M . This turns the scale into one with a *variable origin*. Many examples may be found of preference scales of such a kind. The artist may prefer to work with a BBB pencil, the draughtsman with an HHH; most people prefer a B or an HB. Some epicures may prefer the Tio Pepe or the Manzanilla, and others maintain that there is nothing to equal an old Oloroso, but at a sherry party the medium golden usually disappears the fastest. I have data on women's preferences for lipsticks which exhibit similar phenomena very clearly.

When there is some tendency towards agreement on a single reversible order, but some error variance intrudes, we have no justification for expecting evidence of an agreed order in the mean values of t_{io} in different O -scales. The magnitude and signs of the net differences between the means may depend on how far the balance of opinion tips in favour of the A or M end. If the balance

is even, the resultant order of means may be determined by minor components of the error variance, and the importance of the principal component may be obliterated.

When p can move freely anywhere along the prevalent order of preference, the O -scales may be expected to have equal means and rectangular distributions, like those expected on the assumption of discernment without agreement. The evidence for a prevalent order will be accumulated in the covariances, where a wide range of positive and negative totals will be obtained. Suitable interchanges of rows and columns should collect the highest positive values in the diagonals nearest the leading diagonal and the negative values in the remoter off-diagonals, thus producing a matrix like those for Markoff chains. In the general case, where several different orders of preference are operative, the means on the O -scales will provide only inadequate and often misleading evidence. But if the deviance of every t_{i0} is measured from the general mean of T instead of from the mean of the O -scale where it belongs, all the evidence of agreement is accumulated in the matrix of variances and covariances. The variance between the means is retained as part of the total variance, and can be analysed into the parts contributed by the principal components successively. If a single irreversible order predominates it emerges as the first component.

The matrix required is the O -matrix defined algebraically as $U'U$. Each row of entries in U is now used to specify the co-ordinates of an individual in the O -space, the m -dimensional continuum of preferences for different objects. The O -matrix defines the contours of the region the individuals occupy. Table IV shows the O -matrix obtained from Table II.

TABLE IV. O -MATRIX
defining the dispersion of the individuals in the O -space

Object	Object				
	A	B	C	D	E
A	14	-3	-8	-12	9
B	-3	14	-7	2	-6
C	-8	-7	16	6	-7
D	-12	2	6	18	-14
E	9	-6	-7	-14	18

Principal component analysis can also be applied directly; and whether it is applied to the $m \times m$ O -matrix or the $n \times n$ I -matrix, the quantity finally analysed will be the same, namely, V , the total variance of the t_{i0} about their general mean; for V is the sum of the elements in the leading diagonal both of $U'U$ and of UU' .

V. THE NATURE OF THE DATA USED FOR ILLUSTRATION

The nine objects of which A , B , C , D and E in Table I are five, are solid wooden cylinders, each with a different height and diameter and painted in a different shade of blue. When ranked for height, diameter, or colour, a different

ranking is obtained, independent of the other two ($\tau=0$). Height, diameter, or colour may behave as a reversible scale or as a scale with a variable origin; height and diameter may interact to generate new scales of volume, proportionality, etc.

Many observers who were merely asked to 'compare them two at a time' remarked that they preferred the 'larger' or the 'smaller'; hardly any referred explicitly to height, and none to diameter. The order given by the first component agrees closely with that for volume ($\tau=.78$). Other observers commented on the colour, saying (e.g.) 'I had a preference for royal blue', 'I didn't like them too pale or too dark', 'The closer the colour was to Cambridge blue'. And in fact the order given by the second component agrees closely with that for colour ($\tau=.78$).

Order by volume is reversible: though the majority of those who chose by volume favoured the larger cylinder, a substantial minority preferred the smaller. Colour would provide a scale with a variable origin if the observers' comments were admitted as evidence; but no single irreversible order governs their choices, and the variance between means for different objects forms only a small proportion (8.5 per cent) of V . An analysis of the whole of V (s. sqrs. = 1372) and another of the variance between the means (s. sqrs. = 112) may be placed side by side:

Attributable to	Percentage of the Sum of Squares	
	of the total variance	of the variance between the means
The first component,	46.3	80.0
The second,	23.6	0.0
The remaining six,	30.1	20.0

Thus the contribution of a component to the variance between the means is not necessarily proportional to its contribution to the total variance. But perhaps the most important statistics to compare are the small proportion of V , 8.5 per cent, attributable to the variance between the means, and the substantial proportion, 46.3 per cent, attributable to the first component.

The data in Table I have been extracted from this experiment to exhibit the method of analysis. The order of the cylinders is *C, A, B, D, E* from largest to smallest by volume and *B, A, E, C, D* from darkest to lightest by colour. *I4* and *I6* follow the order by volume exactly in making their choices and *I7* and *I8* follow it closely. But *I4* and *I7* prefer the large cylinders and *I6* and *I8* the smaller. So the net effect of their choices on the column totals is very slight. *I2* and *I5* follow the order by colour closely or exactly, and both prefer the darker. So the effect of their choices on the column totals is more marked.

VI. A LAW OF COMPARATIVE JUDGMENT

If any proposition deserves to be enunciated as a law of *comparative* judgment (13), it is that such judgments are interdependent. When O is presented for comparison with one or more other objects, the observer's response to O is conditioned by the other objects present. Person-object interactions produce the whole variance V . Some implications of this interdependence have already been noted. Another is that preference for one object cannot in general be uncorrelated with preference for another. If J is preferred to K , K cannot be preferred to J , and so the tendency is for preference for one object to be *negatively* correlated with preference for another. Accordingly the matrices obtained by correlating O -scales exhibit a tendency to negative correlation, and Thurstone's method of factor analysis (14) is not strictly appropriate. It may sometimes work out very badly or even give absurd results (8, 9). A principal component analysis may not be completely satisfactory every time; but it usually helps to reveal the nature of the results, and always gives a rational (and as a mathematician would say 'real') solution.

The general tendency can be found by examining the entries in any one row or column of $U'U$ —for example, those of Table IV. They always sum to zero. As the variances in the leading diagonal are necessarily positive, the covariances must be predominantly negative. The total is the covariance of t_{io} with $(t_{ia} + t_{ib} + \dots + t_{im})$, which is the constant $m(m-1)/2$; and the covariance of a variable with a constant is necessarily zero. Centring the variance of each O -scale at its own mean leaves this property unaffected. Accordingly, the matrix of the variances and covariances of the O -scales, each centred at its own mean, may be written as

$$\begin{bmatrix} A^2 & AB r_{ab} & \dots & AM r_{am} \\ AB r_{ab} & B^2 & \dots & BM r_{bm} \\ \dots & \dots & \dots & \dots \\ AM r_{am} & BM r_{bm} & \dots & M^2 \end{bmatrix}$$

where A^2 is the variance on the A -scale, etc. Now let us ask under what conditions the correlations r_{ab} etc. can all be equal. Since the terms in the first row sum to zero we have $A^2 + rAB + \dots + rAM = 0$. Putting $A + B + \dots + M = S$, we may write this as $rA(S - A) = -A^2$. Hence $A = rS/(r-1)$, unless $A = 0$ (i.e. unless there is no variation in preferences). Similarly for the second row. Assuming some variation in preferences, we find that the constant $k = rS/(r-1) = A = B = \dots = M$. So all the variances $A^2, B^2, \dots, M^2$ must be equal. And $S = A + B + \dots + M = mk$. Hence $r k(mk - k) = k^2$; and $r = -1/(m-1)$, unless $k = 0$. This is the only constant value for r compatible with any variation in preferences. If $A = 0$ or $B = 0$, etc., the correlations can only be equal when

$r=0$ and $A=B=\dots=M=0$. This would imply perfect agreement on a single irreversible order. In the remaining cases the correlations must differ.

Thus, when there is any variation between individual's preferences, the assumption that all the correlations are equal, which receives most attention in the literature (3, 7, 15) limits discussion to an extremely restricted range of possible covariance matrices, namely, those where the variances are equal and the correlations are $-1/(m-1)$. This range includes the matrices obtained when there is no discernment and when there is discernment without agreement; and may perhaps be adequate as an approximation in cases where the only evidence sought is for a single irreversible order. But presumably experimenters will be glad to escape from such a restriction if they can. Usually the correlations between the O -scales vary about a central value of $-1/(m-1)$; and, if some are large and positive, others will be negative and even larger. By way of illustration, Table V gives those derived from Table I.

TABLE V. CORRELATIONS BETWEEN THE O -SCALES

Object	B	C	D	E
A	-0.259	-0.544	-0.748	0.577
B		-0.476	0.204	-0.385
C			0.375	-0.412
D				-0.825

The procedures advocated for re-scaling the t_{io} (2, 15) can claim to be appropriate only when certain assumptions are satisfied. There may be good *a priori* grounds for assuming (i) that the variation of observers' preferences for an object, $\mathcal{J}$, tends to be normally distributed about a mean degree of preference, especially if the observers are a random sample drawn from a normally distributed population; (ii) that preferences for some other object, K , will also be normally distributed about a different mean; and (iii) that preferences for $\mathcal{J}$ will tend to be uncorrelated with preferences for K unless there is some psychological connection between them. It may be doubtful whether the preferences for $\mathcal{J}$ vary more widely or less than those for K ; so the assumption of equal variances may be practically defensible. But nothing in any preference table *compels* acceptance of these assumptions. So there may still be occasions when the experimenter feels content to examine a preference table directly, to see what evidence it contains; and in such a case he may welcome a systematic method for mapping his observations. I cannot at present promise that a principal component analysis will help him any further than that.

VII. COMMON PROPERTIES OF THE I -MATRIX AND THE O -MATRIX

The latent roots and vectors of the ellipse defined by a matrix of squares and products M can be obtained by solving the equation

$$Ml = \lambda l$$

where λ is any non-zero latent root and l the corresponding latent vector. Working methods for obtaining the solution are described by Burt (1), Thomson (12), Holzinger and Harman (4), and others.

The latent roots and vectors of the I -matrix UU' are obtained by solving

$$(UU')L = \lambda L; \quad (1)$$

and those for the O -matrix $U'U$ by solving eqn. (2)

$$(U'U)l = \lambda l. \quad (2)$$

UU' is $n \times n$, and L must therefore have n entries; $U'U$ is $m \times m$, so l has m entries. Consequently in general L cannot be identical with l , though their common origin in U suggests that they must be related to one another¹.

If we premultiply eqn. (1) by U' we obtain $U'(UU')L = U'\lambda L$. This can be re-written as $(U'U)U'L = \lambda U'L$, since λ is a scalar. Consequently any latent root, λ , of UU' is also a latent root of $U'U$; and the latent vector associated with it there can be defined as $U'L$. Hence

$$U'L = cl,$$

where c is a constant of proportionality. Further premultiplication, by U , proves that

$$c = \lambda^{\frac{1}{2}}.$$

The same argument applies to eqn. (2). This means that the latent roots and vectors of the I -matrix, which describes the dispersion of the objects in the I -space of individuals' preferences, can be obtained by applying component analysis to the O -matrix, which describes the dispersion of the individuals in the O -space of preferences for objects; and *vice versa*. The more convenient matrix to analyse will be the O -matrix, since in general $m < n$. Though the I -space is n -dimensional it contains only m points; so dispersion in it cannot extend over more than $m-1$ dimensions and the I -matrix cannot have more than $m-1$ non-zero latent roots. The remaining $n-m-1$ must be zero. It follows that the O -matrix also cannot have more than $m-1$ non-zero roots; and one must be zero. Consequently both matrices have zero determinants and no reciprocals. A transformation may be required before techniques of multivariate analysis can be applied to them. Transforming the (t_o) into the L or the Ul for all non-zero λ is usually as good a way as any. When the results of analysing one of the matrices are applied to the other, the constant c keeps intruding. To avoid confusion I would recommend keeping L and l normalized, (*cf.* Burt, 1).

VIII. THE PRINCIPAL COMPONENTS FOR THE ILLUSTRATIVE DATA

The latent roots of the O -matrix in Table IV are given in Table VI with their normalized vectors. The total of the latent roots is V. The four roots account for 57.0, 27.1, 8.1, and 7.8 per cent of V respectively. The first vector places the objects in the order C, A, B, D, E , their order by volume from largest to smallest; and the second in the order B, E, A, C, D , their order by colour

¹ The first to demonstrate this relation in work on factor analysis was Cyril Burt in various papers on 'correlating persons' as compared with correlating tests: *cf. Brit. J. Psychol.*, XXVIII, 1937, pp. 59-96.

from darkest to lightest when E and A are interchanged. So the analysis clearly indicates that the individuals' choices are generally in accordance with the volume or the colour of the cylinders—volume most commonly and colour next—and are also influenced by other considerations to a slighter extent.

TABLE VI. ANALYSIS OF THE O -MATRIX

Object	Latent vectors and latent roots			
	l_1	l_2	l_3	l_4
A	0.475	0.025	0.652	0.385
B	-0.110	0.744	0.030	-0.483
C	0.558	-0.204	-0.666	-0.067
D	-0.350	-0.632	0.247	-0.465
E	-0.573	0.067	-0.264	0.631
Roots	45.6	21.7	6.5	6.2

The dispersion of the individuals in the O -space can be mapped when the co-ordinates giving their location or its principal axes are obtained from V . They are listed in Table VIII. The entries in Table VII serve to define the

TABLE VII. CO-ORDINATES

locating the individuals on the principal axes of the O -space

Individual	Ul_1	Ul_2	Ul_3	Ul_4
1	-1.86	1.46	-2.05	0.47
2	0.40	2.98	0.88	0.42
3	2.08	-1.01	0.33	2.13
4	3.09	0.12	-0.40	-0.55
5	1.53	2.71	0.48	-0.28
6	-3.09	-0.12	0.40	0.55
7	2.85	-1.26	-0.18	-0.53
8	-2.86	-0.82	0.91	-0.55

latent vectors of the I -matrix in Table III. They only need to be normalized and then the analysis in Table VIII can be completed. But these results are

TABLE VIII. ANALYSIS OF THE I -MATRIX

Latent vectors and latent roots

Individual	L_1	L_2	L_3	L_4
1	-0.275	0.313	-0.804	0.189
2	0.059	0.640	0.345	0.169
etc.	...	...	...	...
Roots	45.6	27.7	6.5	6.2

not required for obtaining the co-ordinates of the objects on the principal axes of the I -space. The equivalence $U'L = \lambda^{\frac{1}{2}}l$ permits them to be obtained from Table VI. The co-ordinates locating the objects on the first two principal axes of the I -space may be tabulated as follows:

Object	$\lambda^{\frac{1}{2}}l_1$	$\lambda^{\frac{1}{2}}l_2$
<i>A</i>	3.211	0.117
<i>B</i>	-0.741	3.464
<i>C</i>	3.769	-0.948
<i>D</i>	-2.367	-2.943
<i>E</i>	-3.872	0.311

In some experiments diagrams mapping the dispersions of test-objects and of persons tested may serve to focus attention on some previously unnoticed characteristics of the objects or resemblances between them. But in this experiment so much is known about them already from other sources that such diagrams would add nothing new.

IX. EXPLAINING THE NET RESULTANT ORDER

The preferences expressed by different individuals for a particular object vary about a mean; and the means for the different objects serve to arrange them in an order. This is the net resultant of all the person-object interactions. When the individuals' preferences tend to conform with a single irreversible order, the means will do so too. A preference table exhibiting this sort of agreement can be analysed by the method described above. The predominant order will be defined by the vector for the largest of the principal components. Its latent root will account for a large proportion of V , and the roots of the remaining components will be comparatively small, so the points locating the objects in the I -space will tend to converge on a straight line. This will be the best solution as defined by the method of least squares. Under these conditions, as is well known, the means for the objects may themselves give a very good approximation to the first latent vector, and the sum of squares due to the differences between them may be almost as large as the first latent root. Thus in such a case the experimenter may be able to extract all the information required simply by computing the means and their variance.

When a reversible order or several orders of preference are operative, none is likely to monopolize the variance between the means. No useful approximation to any of the latent vectors may be obtainable from the means, or they may approximate better to one of the later vectors rather than to the first. But their variance can always be expressed as the resultant of effects due to all the components.

In Table IX the first column of entries denoted e_1 , is provided by the column totals of U in Table II. It shows the net resultant order with totals replacing means. The two subsequent entries are $e_1'e_1/n$ (the sum of squares accounted

for by the net resultant order) and $e_1'l=w_1$, which should be positive. The second column, $e_2=w_1l_1$, is the set of estimates which would fit e_1 if the variance between the means were attributable to the first of the principal components. The third column, $e_3=e_1-e_2$, is the set of differences from expectation. Every column total is necessarily zero (a useful check); and $e_3'e_3/n$ is the residual sum of squares not attributable to the first component. The vector of the second component is now used to obtain: $e_3'l_2=w_2$; hence the entries in the fourth column, $e_4=w_2l_2$; and in the fifth, $e_5=e_3-e_4$. And so on, until each latent vector has been used.

TABLE IX. ANALYSIS OF THE VARIANCE BETWEEN THE MEANS

Object	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
<i>A</i>	2	1.013	0.987	0.102	0.885	0.244	0.640	0.640
<i>B</i>	2	-0.234	2.234	3.027	-0.793	0.011	-0.804	-0.804
<i>C</i>	0	1.189	-1.189	-0.828	-0.361	-0.249	-0.111	-0.111
<i>D</i>	-4	-0.747	-3.253	-2.572	-0.682	0.093	-0.774	-0.774
<i>E</i>	0	-1.222	1.222	0.271	0.951	-0.099	1.049	1.049
Total	0	0	0	0	0	0	0	0
$e'e/n$	3.0		2.432		0.364		0.346	
$e'l$	2.132		4.067		0.375		1.664	

The differences between the successive values of $e'e/n$ provide an analysis of the sum of squares of the variance between the means. The results in Table IX indicate the following proportions for the four components: 18.9, 68.9, 0.6, and 11.5 respectively. The figures corroborate the conclusions drawn from the preliminary inspection of the data. Preferences in colour favour the darker blues, as already noted; and accordingly the component associated with colour accounts for the largest part of the variance between the means. The first component, associated with volume, accounts for a comparatively small proportion of the variance. But the whole sum of squares is small when compared with the total variance due to person-object interactions, viz. 3/80 or 3.75 per cent.

X. GENERAL CONCLUSIONS

At present these statistics should be regarded as descriptive. They lead to no tests of significance or support no inferences from sample to the population. The conclusions drawn from the data in Table I do not entirely coincide with the conclusions reached in the experiment from which the data were extracted; and the results of a larger experiment might entail further modifications. I hope, however, that the potentialities of the method may appear attractive enough to encourage investigations into the application of probability theory to the problems involved.

The general properties of the dispersion of the *I*-points locating individuals

in the O -space are worth considering from this point of view. All P_i with $i=0$ have $v_{ix}=\phi_{im}$, which is a constant; and all P_i with $i>0$ have $v_{ix}<\phi_{im}$. Hence the dispersion has a spherical boundary at a distance $\sqrt{(\phi_{im})}$ from the multivariate mean point. Suppose x is small and m large. Then, assuming no discernment, the density of the I -points will tend to be equal at all points equidistant from the mean. There will be a region of maximum density at a certain distance from the mean, and thereafter the density will diminish steadily as the distance approaches $\sqrt{(\phi_{im})}$. The dispersion will accordingly form a hypersphere with a certain solidity, least nearest its surface. Assuming discernment without agreement the I -points will disperse away from the mean towards the boundary evenly in all directions, and tend to form a solid envelope around an empty hypersphere. Assuming agreement (and discernment implicitly) the I -points will be distributed over the hypersurface, but not evenly; they will collect in certain areas just as the human population of this planet collects in certain areas on its surface and leaves other areas comparatively empty. Thus we have three neat models for representing no discernment, discernment without agreement, and agreement. Of course a further elaboration of models for representing various kinds of agreement could be contemplated.

All kinds of agreement can be represented graphically by making suitable use of results obtained by this method. It might, for instance, be employed commercially to investigate the density of the population in psychological areas of unsatisfied demand. The argument of selection VII that an $n \times n$ covariance matrix can always be replaced by an $m \times m$ one for the purposes of a principal component analysis is quite general. As Burt has insisted, it can be applied to any table of scores of n persons on m tests. Thus it demonstrates a point of convergence of P and R techniques (1, 11). It is a restatement in a more general form of the relation described by Burt. But the tables of test scores he was considering—say S , with its characteristic element s_{io} denoting the score of the i th person on the o th test—are less tractable than preference tables. The s_{io} in a row of S are not initially commensurable. Burt apparently did not consider normalization by row and by column as two alternative possibilities. In his cases where he used the double technique he assumed the s_{io} would have to be normalized both by row and by column before equivalent results could be obtained from an $n \times n$ matrix of correlations between persons and an $m \times m$ matrix of correlations between tests. Simultaneous normalization by row and by column is possible only under exceptional conditions; and accordingly his argument was in fact criticized by Thomson (12) on the ground that it lacks generality. Though an $m \times m$ matrix can always be constructed to replace an $n \times n$ matrix, or vice versa, the results of the analysis are modified when normalization by row replaces normalization by column. P and R techniques diverge from one another towards different kinds of experimental design, and thus produce data with different specifications. But the same statistical methods may apply to the data obtained from either source, or indeed from any other source; the limitations to their applicability are formal.

APPENDIX

A MERCURY PROGRAMME

Mr. W. L. B. Nixon of the University of London Computer Unit has developed a programme for the Mercury computer to carry out analyses by the method described in this paper. The programme (Preference Analysis, ref. S9.1/WLBN/ULCU) accepts data in the form of an $n \times m$ preference table, T , such as the one in Table I. The number of objects, m , must not exceed 28, but n , the number of individuals, is practically unlimited. The data are centred by subtracting $(m-1)/2$ from each element. The programme then computes and prints out (i) the m column-sums of the resultant U matrix, as shown in e_1 , Table IX, and (ii) the $m \times m$ symmetrical dispersion matrix $O = U'U$, omitting elements above the leading diagonal. Next the programme computes the latent roots of the O -matrix in succession in order of magnitude, and prints them together with their associated latent vectors, normalized, thus specifying the principal axes of the O -dispersions. The machine operator can terminate this operation after any given number of the roots and vectors have been extracted, if all are not required. Finally the programme re-reads the original, or a similar set of data and prints out a set of co-ordinates Ul , such as those in Table VII, for each individual, locating him with respect to the major axes of the O dispersion. This operation need not extend to all the axes previously computed, but can be restricted by the machine operator to a smaller number if desired. The time of execution of the programme increases approximately linearly with the value of n and with the cube of the value of m . Typical total times vary from a few minutes for small values of m to half-an-hour or so for values of m near 28, when the samples are of a moderate size, say n 200.

REFERENCES

- [1] BURT, C. (1940). *The factors of the mind*. Macmillan.
- [2] FISHER, R. A. and YATES, F. (1943). *Statistical tables for biological, agricultural and medical research*. Oliver and Boyd.
- [3] GUILFORD, J. P. (1954). *Psychometric methods*. New York : McGraw-Hill.
- [4] HOLZINGER, K. J. and HARMAN, H. H. (1954). *Factor analysis*. Chicago : University Press.
- [5] KENDALL, M. G. and BABINGTON SMITH, B. (1939). On the method of paired comparisons. *Biometrika*, XXXI, 324.
- [6] KENDALL, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, XI, 43.
- [7] MOSTELIER, F. (1951). Remarks on the method of paired comparisons. *Psychometrika*, XVI, 203.
- [8] SLATER, P. (1951). The transformation of a matrix of negative correlations. *Brit. J. stat. Psychol.*, IV, 9.
- [9] SLATER, P. (1953). The factor analysis of matrices of negative correlations. *Brit. J. stat. Psychol.*, VI, 101.
- [10] SLATER, P. (1960). *The reliability of some methods of multiple comparison in psychological experiments*. Ph.D. Thesis, University of London.
- [11] STEPHENSON, W. (1955). *The study of behavior*. Chicago : University Press.
- [12] THOMSON, G. (1951). *The factorial analysis of human ability*. London : University Press.
- [13] THURSTONE, L. L. (1927). A law of comparative judgment. *Psychol. Rev.*, XXXIV, 273.
- [14] THURSTONE, L. L. (1947). *Multiple factor analysis*. Chicago : University Press.
- [15] TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York : John Wiley & Sons.
- [16] WEI, T. H. (1952). *The algebraic foundations of ranking theory*. Thesis, University of Cambridge.

PATTERN ANALYSIS¹

By H. J. A. RIMOLDI and T. F. GRIB

Loyola University, Chicago, Illinois

The purpose of the following article is to explain a technique for describing and comparing bivariate patterns in which a number of subjects respond to a number of items. A new scoring procedure is proposed for the responses based on a simple weighting principle, which leads to a convenient formula for comparing two or more patterns. The advantages of the method are discussed; and it is suggested that it may have useful applications in other fields.

I. THE DESCRIPTION OF PATTERNS

A variety of techniques for analysing patterns have recently been published. The chief procedures have been summarized in [2, 6, 7]. Of these some analyse the distances between stimuli [4, 1], others use the chi-square technique [8], or the phi coefficient [7]. Guttman [5] has described his concept of reproducibility: according to his 'scalogram hypothesis', "the items have an order such that ideally persons who answer a given question favourably all have higher ranks on the scale than persons who answer the same question unfavourably". Thus in the ideal situation, if the inter-item covariances are maximal and we know the respondent's rank or scale score, it is possible to deduce the number of items he endorsed. The index of consistency developed by Green [3] has advantages over those previously described. But neither his nor Guttman's technique provide a way of estimating the reproducibility of a single item. This difficulty was surmounted by both J. M. Jackson and J. Loevinger [7].

Given a set of n stimuli, and provided we know how subjects behave in relation to one or more properties of the stimuli, it is possible to define a pattern of responses for each item or each subject or the whole group of subjects and items. The term stimulus is here understood in its broadest sense, and includes any test item, test, physical object, verbal statement, or an item of behaviour. The subject's responses to each of the stimuli may be classified under one or other of two mutually exclusive categories, such as passing or not passing a test item, having a score above or below a predetermined cutting point, selecting or rejecting a given object or item of behaviour, or agreeing or disagreeing with a verbal statement².

¹ This problem was approached as a question so-called 'run theory' by the authors and J. R. Devane. The cumbersomeness of the calculations gave rise to this investigation, and led to the development of the present technique. The authors wish to thank Dr. H. P. Bechtold for his comments and suggestions. This project was partly supported by a grant from the Commonwealth Fund of New York.

² Although in this paper we are dealing with dichotomous responses, the procedure can be extended to responses to stimuli that may be classified into n mutually exclusive categories.

Let us assume that a group of five subjects state their choice among eight school courses as indicated in Figure 1. In this figure each subject may be characterized by his peculiar arrangement of filled and empty cells. This arrangement constitutes what we define as an individual row pattern. The arrangement for the entire table will be called the total pattern. Patterns may also be described for each stimulus (i.e., down the columns of Figure 1) or for the whole sample of stimuli. According to the problem, the investigator may be interested in describing patterns of subjects or patterns of items. In either case the method to be applied will be essentially the same.

FIGURE 1. PATTERN OF RESPONSES

Subjects	School Courses								R_i	$\bar{R}_i$
	A	B	C	D	E	F	G	H		
1	X		X			X	X		4	4
2		X		X				X	3	5
3	X			X		X		X	4	4
4			X		X				2	6
5				X					1	7
C_j	2	1	2	3	1	2	1	2	14	—
$\bar{C}_j$	3	4	3	2	4	3	4	3	—	26

Note.—X indicates that the subject chose the course designated at the top of the columns.

For purposes of nomenclature we shall designate each filled-in cell a_{ij} , where, following the usual matrix notation, i indicates rows and j columns. Let $\bar{a}_{ij}$ denote the empty cell in row i and column j . Obviously the sum of all the a_{ij} and $\bar{a}_{ij}$ cells along any row (subjects) will be equal to n (the total number of stimuli); and the sum of all the a_{ij} and $\bar{a}_{ij}$ cells along any column (stimuli) will be equal to k (the total number of subjects). For each row the sum of a_{ij} will be called R_i and the sum of all $\bar{a}_{ij}$ $\bar{R}_i$. Similarly for each column, the sum of all a_{ij} cells will be called C_j and the sum of $\bar{a}_{ij}$ cells $\bar{C}_j$. The value of R_i for different subjects may be the same in spite of the fact that the location of the a_{ij} and $\bar{a}_{ij}$ cells is not the same (subjects 1 and 3 in Figure 1). Similarly for $\bar{R}_i$, C_j , and $\bar{C}_j$. Therefore the R_i and $\bar{R}_i$ values for a subject are a necessary but not a sufficient condition for characterizing his pattern.

A system of weights can be defined in terms of the designation of a cell as the intersection of a row and a column, such that the total contribution of the corresponding arrays is taken into account. Accordingly it will be assumed that a filled or empty cell at the intersection of two arrays with high marginal values contributes more to the quantitative characterization of a pattern than a cell at the intersection of two arrays with low marginal values. Thus for all a_{ij} cells the weight (W) will be defined as

$$Wa_{ij} = \frac{R_i C_j}{\sum_i R_i}; \quad (1)$$

and similarly for all the $\bar{a}_{ij}$ cells,

$$W\bar{a}_{ij} = \frac{\bar{R}_i \bar{C}_j}{\sum_i \bar{R}_i}. \quad (2)$$

In terms of eqns. (1) and (2) the weight to be assigned to each cell is a function of the product of R_i and C_j , and of the total number of a_{ij} cells. Since a_{ij} and $\bar{a}_{ij}$ cells represent the presence or absence of a particular attribute or trait, both should be considered in characterizing the pattern. Thus according to eqn. (1), cell a_{2B} in Figure 1 has a weight of $(3 \times 1)/14 = 0.21$; and cell $\bar{a}_{3C}$ a weight of $(4 \times 3)/26 = 0.46$.

In some cases different cells may contribute the same amount (e.g., cells a_{1A} , a_{1C} , and a_{1F} in Figure 1) and so will have the same weight. Thus, if two like cells (a_{ij} or $\bar{a}_{ij}$) occupy positions having the same marginal totals, they will have the same weight. In general, if item a has been selected by k subjects, and if item b has been selected by the same k subjects, then in terms of our weights it will not be possible to differentiate item a from item b , and we shall call them equivalent. The same holds good for the response pattern of subjects: i.e., if two or more subjects select the same items, their response patterns will be equivalent, and so will the values of R_i , $\bar{R}_i$ and the weights of all cells in the corresponding pattern.

The sum of all the weights in the whole table, including both a_{ij} and $\bar{a}_{ij}$ cells, is the total, T . The sum of weights along each row gives the total for each row (subject), T_{ri} ; and similarly a total for each column (stimulus), T_{cj} , can be calculated. While T characterizes the whole table, T_{ri} characterizes row i and T_{cj} column j . In the following text the subscript r and c designate rows and columns respectively, and i and j specific rows or columns.

The above method of assigning weights may be conveniently used when the responses are dichotomous and no *a priori* assumptions as to the pattern are expected. The weights are determined by the performance of the group itself. Occasions may arise, however, when other methods of assigning weights would be more advantageous. In the case of non-dichotomous responses (i.e., when the response is represented by a score indicating amount rather than the mere presence or absence of a trait or type of behaviour) the person's score for each item may also be used for assigning weights.

II. THE COMPARISON OF PATTERNS

Various indexes can be developed which will enable us to compare either individual patterns or patterns for a whole table with an expected 'model' pattern or another observed pattern. Let us assume that in a given experiment

the results obtained were as shown in Figure 2. Our problem is to investigate how this observed pattern differs from an expected pattern, under the condition that the total scores for each individual will be the same. Assume that the expected pattern is as shown in Figure 3 a; and calculate for this model (according to formulae (1) and (2)) the corresponding weights¹. The results of these operations are given in Figure 3 b.

The similarity between the two patterns, model and observed, may be expressed by the following ratio:

$$\text{Similarity} = \frac{\text{sum of weights of unchanged cells}}{\text{total weight of model pattern}}.$$

The similarity ratio between the patterns in Figure 2 and Figure 3 a may then be estimated by adding for each row in succession the weights for the cells that have not changed. Call these values A_i , and their sum A_t . (Note that when, in any row, all cells are of the a_{ij} type, $A_i = R_i$, and when they are of the $\bar{a}_{ij}$ type, $A_i = \bar{R}_i$.) In the first row (Figure 2 and 3), cells a_{1A} and a_{1B} remain unchanged; so the value of A_1 is $1.50 = (.90 + .60)$. In the second, the cells that remain unchanged are $\bar{a}_{2A}$ and a_{2C} , and their sum (A_2) will be $(.33 + .60) = .93$. And so on. Summing all these A_i values gives $A_t = 7.93$, which is a statement concerning cells that are the same in the observed and model patterns. These operations imply comparing two matrices, adding weights of like elements, and dividing by the total weights of the model pattern. The similarity ratio (${}_iS_r$) between the observed pattern (Figure 2) and the model pattern (Figure 3 a) may be defined as

$${}_iS_r = \frac{A_t}{T}. \quad (3)$$

FIGURE 2. OBSERVED RESPONSE PATTERN

Subjects	Test Items				R_i	$\bar{R}_i$
	a	b	c	d		
1	X	X		X	3	1
2		X	X		2	2
3	X	X	X	X	4	0
4				X	1	3
C_j	2	3	2	3	10	—
$\bar{C}_j$	2	1	2	1	—	6

¹ We may also have an *a priori* definition for the expected pattern. In this case we may build it independently of the performance of any group, and use it as a model against which the group's performance is compared. In this case the weights may be arbitrarily assigned on the basis of what we expect *should happen*, not on the basis of actual performance.

FIGURE 3 a. MODEL PATTERN

Subjects	Items				R_i	$\bar{R}_i$
	a	b	c	d		
1	X	X	X		3	1
2			X	X	2	2
3	X	X	X	X	4	0
4	X				1	3
C_j	3	2	3	2	10	
$\bar{C}_j$	1	2	1	2		6

FIGURE 3 b. MATRIX OF WEIGHTS FOR MODEL PATTERN

Subjects	Items				T_{ri}	A_i	${}_mA_i$
	a	b	c	d			
1	.90	.60	(.90)	(.33)	2.73	1.50	1.50
2	.33	(.67)	.60	(.40)	2.00	.93	.00
3	1.20	.80	1.20	.80	4.00	4.00	4.00
4	(.30)	1.00	.50	(1.00)	2.80	1.50	1.50
					$T=11.53 \ A_t=7.93 \ {}_mA_t=7$		

Note.—The values in parenthesis indicate the cells which differ in the model and the observed pattern. The values in bold figures correspond to $\bar{a}_{ij}$ cells in the model.

In our experiment its value will be $7.93/11.53=0.69$. If the model and the observed pattern are identical, then ${}_iS_r=1.00$, since in this case $A_i=T$. The difference between A_i and T is an indication of the amount of deviation between the patterns compared. If all the a_{ij} cells become $\bar{a}_{ij}$ cells and vice versa, the value of A_i becomes zero; so does ${}_iS_r$. But this will, in most cases, violate the condition of keeping the values of R_i and $\bar{R}_i$ constant.

A minimum value should be defined for each pattern such that the final similarity ratio will be equal to zero when the observed pattern shows the greatest deviation from the expected pattern. The index of agreement (I_a) for any table, under any conditions (except when all the cells are of the a_{ij} type or $\bar{a}_{ij}$ type) will then vary from 1.00 (perfect agreement) to 0 (minimum possible agreement). The pattern showing the greatest deviation will be the one that, keeping the row marginals (R_i and $\bar{R}_i$) constant, will have, in relation to the model pattern, the minimum possible number of a_{ij} and $\bar{a}_{ij}$ cells in the same location. Further, the weights corresponding to these cells should be the smallest in the row, so that their sum will be a minimum.

Every misplacement implies that an a_{ij} cell in the observed pattern becomes an $\bar{a}_{ij}$ cell in the model, and vice-versa. If, in a given row, the number of a_{ij}

and $\bar{a}_{ij}$ cells is the same, then every cell can be misplaced; but if the number is different, then the number of misplacements is limited by the number of a_{ij} or $\bar{a}_{ij}$ cells, whichever is smaller: i.e., the number of unchanged cells will be $R_i - \bar{R}_i$. This value gives the number of unchanged cells under conditions of greatest deviation. At this point, a convention may be introduced. If for a given row $(R_i - \bar{R}_i)$ is positive, then the non-misplaced cells are of the a_{ij} type; but if it is negative, then the non-misplaced cells are $\bar{a}_{ij}$.

TABLE 1. NUMBER OF UNCHANGED CELLS
VALUES FROM FIGURE 3

Row	$(R_i - \bar{R}_i)$	Weights	${}_m A_i$
1	2	(.60 + .90)	1.50
2	0	—	.00
3	4	(1.20 + .80 + 1.20 + .80)	4.00
4	-2	(.50 + 1.00)	1.50
			7.00 = ${}_m A_t = \sum {}_m A_i$

Since, for each row, the sum of weights (${}_m A_i$) has to be a minimum (to determine minimum possible agreement), after finding the number and type of non-misplaced cells, we add the weights of those with the lowest values; this gives the smallest value for ${}_m A_i$: e.g., in row 1 in Figure 3 there can be only two a_{ij} non-misplaced cells; and, to fulfil the condition of a minimum sum, those to be added must therefore be .60 and .90. In row 2 there can be no non-misplaced cells, i.e. the sum will be 0, but in row 3 none can be misplaced, and their sum is 4.00. Thus the total value (${}_m A_t$) is 7.00.

${}_m A_i$ and ${}_m A_t$ indicate respectively the sums of the weights of the minimum possible number of non-misplaced cells in relation to the model by rows and for the whole table. The letter m indicates 'minimum value'. When the cells in a row, column, or table are all of the a_{ij} type or the $\bar{a}_{ij}$ type, there will be no minimum. The similarity ratio under the condition of greatest deviation will be, according to (3),

$${}_m S_a = \frac{{}_m A_t}{T}, \quad (4)$$

i.e. in our example $7.00/11.53 = 0.61$.

In order to obtain an index with the value of 0, and so to compare tables of different sizes, an index of agreement (I_a) may be defined by the expression

$$I_a = \frac{{}_t S_r - {}_m S_a}{1.00 - {}_m S_a}. \quad (5)$$

A simpler expression for I_a can be obtained by substituting in (5) the values of ${}_iS_r$ and ${}_mS_a$ obtained in (3) and (4). The equation for I_a then becomes

$$I_a = \frac{A_t - {}_m A_t}{T - {}_m A_t}. \quad (6)$$

When *all* cells are of the a_{ij} type or of the $\bar{a}_{ij}$ type, $A_t = {}_m A_t = T$; and I_a is indeterminate. Any system of cells, keeping row totals constant, must have a possible maximum and a possible minimum: otherwise the system cannot change, and differentiation between patterns is impossible. In the case of the most deviant pattern, A_t in eqn. (6) will be equal to ${}_m A_t$. Then we can write $I_a = (7.00 - 7.00)/(11.53 - 7.00) = 0$. When the model and observed patterns are equal, A_t will be equal to T , and the value of I_a will be $(11.53 - 7.00)/(11.53 - 7.00) = 1.00$. Finally for the observed pattern in Figure 2, A_t is 7.93, ${}_m A_t$ is 7.00, and $I_a = (7.93 - 7.00)/(11.53 - 7.00) = 0.21$. Note that by using eqn. (5) we obtain the same value for the observed pattern in Figure 2, viz. $I_a = (0.69 - 0.61)/(1.00 - 0.61) = 0.21$. Since this index has an absolute maximum of 1.00 and a minimum of 0, independent of the size of the matrix, indexes of agreement for tables of different sizes may be validly compared.

This index may also be adapted to compare either subjects (rows) or items (columns). The formula for subjects is,

$$I_{ai} = \frac{A_i - {}_m A_i}{T_{ri} - {}_m A_i}, \quad (7)$$

and for items

$$I_{bj} = \frac{B_j - {}_m B_j}{T_{cj} - {}_m B_j}, \quad (8)$$

where

- I_{bj} = index of agreement for column j ;
- I_{ai} = index of agreement for row i ;
- B_j = sum of weights of non-misplaced cells in column j ;
- ${}_m B_j$ = sum of weights under most deviant conditions for column j ;
- T_{cj} = sum of weights of all cells in column j ;
- A_i = sum of weights of non-misplaced cells in row i ;
- ${}_m A_i$ = sum of weights under most deviant conditions for row i ;
- T_{ri} = sum of weights of all cells in row i .

For the total table, when working by columns, eqn. (6) becomes

$$I_b = \frac{B_t - {}_m B_t}{T - {}_m B_t}, \quad (9)$$

where B_t = sum of B_j 's, and ${}_m B_t$ = sum of ${}_m B_j$'s.

We can analyse the results for each row by means of eqn. (7): cf. Table II.

TABLE II. ANALYSIS BY ROWS OF MODEL AND OBSERVED PATTERNS

Row	T_{ri}	A_i	$S_{ri} = \frac{A_i}{T_{ri}}$	${}_m A_i$	$I_{ai} = \frac{A_i - {}_m A_i}{T_{ri} - {}_m A_i}$
1	2.73	1.50	.55	1.50	$I_{a1} = \frac{0}{1.23} = 0$
2	2.00	.93	.47	.00	$I_{a2} = \frac{.93}{2.00} = .47$
3	4.00	4.00	1.00	4.00	$I_{a3} = \frac{0}{0} = (\text{indeterminate})$
4	2.80	1.50	1.54	1.50	$I_{a4} = \frac{0}{1.30} = 0$
TOTALS	$T = 11.53$	$A_t = 7.93$		${}_m A_t = 7.00$	$I_a = .21$

Note.—The subscript r indicates that we are dealing with rows.

The values of S_{ri} need not be computed to obtain I_{ai} ; and, although the values of S_{ri} for subjects 1 and 4 are higher than for subject 2, their I_{ai} values are lower. This is because the patterns for subject 1 and 4 have the greatest possible deviation. Hence $I_{ai} = 0$.

The index of agreement for a particular table will vary, according to whether we work by rows (I_a) or columns (I_b). For example, suppose we are interested in analysing the item patterns rather than the subjects' response patterns in Figure 4 a: then we must compare the patterns with respect to a hypothetical or expected pattern. This must fulfil the condition that the total scores (C_j and $\bar{C}_j$) for each item are the same. Assume that the model pattern is as indicated in Figure 4 b; and calculate the corresponding weights (Figure 4 c). The results of the analysis by columns will then be as shown in Table III. Thus, the technique described in this paper may be used for comparisons between individuals or between items, and so may be used for item selection.

Once we have decided on the pattern of responses for each item (either individually or collectively), we can set up a model pattern following the procedure described above. The method of analysis will depend on what information is desired. If we wish to determine the acceptability of each item separately, we can use eqn. (8). To decide which items are suitable and which unsuitable, a cutting point can be fixed and the reasons for its device explained. For the pattern given in Figure 3 the values of possible inversions are listed in Table IV. For instance, if columns b and c are inverted, we have for the successive rows the values 2.73, .73, 4.00, 2.80; and $A_t = 10.26$. This will be the pattern a c b d. The table does not give values for all variations, but only for inversions of whole columns. A test of significance for the index of agreement here proposed will be given in a later paper.

FIGURE 4

c. MATRIX OF WEIGHTS FOR MODEL PATTERN

b. MODEL PATTERN

a. OBSERVED PATTERN

Items							Items							Items						
(S)	a	b	c	d	R	$\bar{R}$	(S)	a	b	c	d	R	$\bar{R}$	(S)	a	b	c	d		
1	X	X		X	3	1	1	X		X		2	2	1	40	(33)	(40)	(33)		
2		X	X		2	2	2		X		X	2	2	2	67	60	(67)	(60)		
3	X	X	X	X	4	0	3	X	X		X	3	1	3	60	90	(33)	90		
4				X	1	3	4		X	X	X	3	1	4	33	(90)	(60)	90		
C_j	2	3	2	3	10		C_j	2	3	2	3	10		T_{ij}	2.00	2.73	2.00	2.73		
$\bar{C}_j$	2	1	2	1		6	$\bar{C}_j$	2	1	2	1	6		B_j	2.00	1.50	0	1.80		
														mB_j	.00	1.50	.00	1.50		
																		$T=9.46$		
																		$B_i=5.30$		
																		$mB_i=3.00$		

Note.—Values in parenthesis indicate the cells which differ in the model and observed patterns. Figures in bold are for $\bar{a}_{ij}$ cells.

TABLE III. ANALYSIS BY COLUMNS OF MODEL AND OBSERVED PATTERNS

Column	T_{cj}	B_j	S_c	mB	I_{bj}
a	2.00	2.00	1.00	0	$I_{ba}=1.00$
b	2.73	1.50	.55	1.50	$I_{bb}=0$
c	2.00	0	0	0	$I_{bc}=0$
d	2.73	1.80	.66	1.50	$I_{bd}=.24$
TOTAL	$T=9.46$	$B_t=5.30$	${}_tS_c=.56$	$mB_t=3.00$	$I_b=.36$

TABLE IV. VALUES FOR SUCCESSIVE INVERSIONS OF COLUMNS IN FIGURE 3

Order of Columns	A_t	${}_tS_r$	I_a
abcd	11.53	1.00	1.00
abdc	10.30	.89	.73
acbd	10.26	.89	.72
acdb	9.23	.80	.49
adbc	9.33	.81	.51
adcb	9.53	.83	.56
bacd	10.23	.89	.71
badc	9.00	.78	.44
bcad	9.46	.82	.54
bcda	7.93	.69	.21
bdca	8.23	.71	.27
bdac	8.53	.74	.34
cabd	9.30	.81	.51
cadb	8.27	.72	.28
cbad	9.80	.85	.62
cbda	8.27	.72	.28
cdab	7.80	.68	.18
cdba	7.30	.63	.07
dabc	8.07	.70	.24
dacb	8.27	.72	.28
dbca	8.27	.72	.28
dbac	8.57	.74	.35
dcab	7.50	.65	.11
dcba	7.00	.61	0.00

Note.—For successive inversions of columns A_t , ${}_tS_r$, and I_a will vary ; T (11.53), mA_t (7.00), and mS_a remain constant.

III. THE ADVANTAGES OF THE TECHNIQUE PROPOSED

In most of the techniques so far available [3, 5, 7] the interest has been to find out whether a particular arrangement of items can be interpreted as perfectly scalable. This implies comparison with a model derived from the data under

the conditions of perfect scalability, and requires the responses to be dichotomized. The operations to be performed imply a rank ordering of the data in terms of the marginals of both rows and columns. In the case of ties (i.e. rows or columns with equal totals) this introduces ambiguity.

The final index of reproducibility is based on a count of errors to which is added in Jackson's solution [7] the estimation of minimum reproducibility. The minimum reproducibility is a function of the number of right *or* wrong answers, whichever is larger. Our approach defines a pattern as a specific arrangement of filled and empty cells, and both are considered in finding the index of agreement: after all, failure to respond is as significant as an actual response. We assume that the contribution of a given cell to a pattern is a function of its marginals. Thus each cell has a differential weight determined by its popularity in terms of both subjects and items and according as it is an a_{ij} or an $\bar{a}_{ij}$ cell. This does not require a previous ordering of items or subjects nor dichotomous responses. Our method implies the subtraction of patterns, model and observed, without regard to the peculiar arrangement of filled or empty cells. The restriction that the marginal totals should be the same, is also implicit in Guttman's and Green's procedures [3, 5]. In our case a perfect scalogram or any other arrangement can be considered as a possible model that may be used. Our method provides an index of agreement (1) for the whole table, keeping rows constant, (2) for the whole table keeping columns constant, (3) for each row, and (4) for each column. The definition of a most deviant pattern makes it possible for any table to obtain indexes of agreement varying between 0 and 1.00. In some cases, however, it may be indeterminate.

FIGURE 5. OBSERVED PATTERNS

a	b	c	d
XXXX	XXXX	XXXX	XXXX
XOXX	XXOX	XXXO	OXXX
OXXO	OXXO	OXXO	OXOX
OOXX	OOXX	OOXX	OOXX

FIGURE 6. MODEL PATTERN

Subjects	a	b	c	d
1	X	X	X	X
2	O	X	X	X
3	O	O	X	X
4	O	O	O	X

The index of reproducibility as defined by Green gives for the observed patterns in Figure 5 a value of .937 when compared with the model pattern of Figure 6: i.e. it does not differentiate between the two. Using the method here described the indexes of agreement are .83, .79, .75, and .81 for patterns a, b, c, and d respectively. The discriminative power of our method is due to the differential characterization of each cell in terms of their weights. Those having the same weights are equivalent; and consequently our procedure will not differentiate between them, but the fact that the a_{ij} and $\bar{a}_{ij}$ cells are not weighted 1 and 0, as in other procedures, makes a greater discrimination possible.

One criticism often leveled against indexes of reproducibility is that they tend to give high values. In Table V the results obtained by using our approach

to estimate the index of agreement obtained by interchanging all the columns in Figure 6 are presented. In every case our values are smaller than those obtained with the method described by Green, and moreover vary between 0.00 and 1.00.

TABLE V. A COMPARISON OF RESULTS OBTAINED WITH THE PRESENT METHOD AND WITH GREEN'S

Order of Columns	A_t	${}_tS_r$	I_a	Green's Method
abcd	13.67	1.00	1.00	1.00
abdc	12.77	.93	.87	.94
acbd	12.40	.91	.81	.94
acdb	11.30	.83	.64	.88
adbc	11.00	.80	.60	.88
adcb	10.80	.79	.57	.88
bacd	12.57	.92	.84	.94
badc	11.67	.85	.70	.88
bcad	11.00	.80	.60	.88
bcda	9.60	.70	.39	.81
bdca	9.10	.67	.31	.81
bdac	9.60	.70	.39	.81
cabd	10.97	.80	.60	.88
cadb	9.87	.72	.43	.75
cbad	10.67	.78	.55	.88
cbda	9.27	.68	.34	.75
cdab	7.80	.57	.12	.75
cdba	7.50	.55	.07	.75
dabc	9.07	.66	.31	.81
dacb	8.80	.65	.27	.75
dbca	8.27	.60	.19	.75
dbac	8.97	.64	.30	.81
dcab	7.30	.53	.04	.75
dcba	7.00	.51	0.00	.75

The procedure described in this article refers to the case of two variables only. A generalized procedure for patterns involving more than two variables has been obtained, and will be published later. We then hope to show how the scoring system here described may be used to reformulate classical concepts in test theory, factor analysis, and scaling methods.

REFERENCES

- [1] CRONBACH, LEE J. and GLESER, GOLDINE C. (1953). Assessing similarity between profiles. *Psychol. Bull.*, L, 456-473.
- [2] GAIER, EUGENE L. and LEE, MARILYN C. (1953). Pattern analysis: The configural approach to predictive measurement. *Psychol. Bull.*, L, 140-148.
- [3] GREEN, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrika*, XXI, 79-88.

- [4] OSGOOD, C. E. and SUCI, G. J. (1952). A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, XLIX, 251-262.
- [5] STOUFFER, S. A., GUTTMAN, LOUIS, SUCHMAN, E. A., LAZARSFELD, PAUL F., STAR, SHIRLEY A. and CLAUSEN, J. A. (1950). *Measurement and Prediction*, IV. Princeton: Princeton University Press.
- [6] TORGERSON, WARREN S. (1958). *Theory and Methods of Scaling*. New York: Wiley and Sons, Inc.
- [7] WHITE, BENJAMIN W. and SALTZ, ELI (1957). Measurement of reproducibility. *Psychol. Bull.*, LIV, 81-99.
- [8] ZUBIN, J. (1936). A Technique for Pattern Analysis. *Psychol. Bull.*, XXXIII, 773.
- [9] ZUBIN, J. (1938). The determination of response patterns in personality adjustment inventories. *J. Educ. Psychol.*, XXVIII, 401-413.

COMPONENTS OF VARIANCE ANALYSIS AS A STRUCTURAL AND DISCRIMINAL ANALYSIS FOR PSYCHOLOGICAL TESTS

By R. DARRELL BOCK

University of North Carolina

An approach to the study of the covariance structure and the discriminating properties of psychological tests is proposed for cases in which the tests can be classified hierarchically and suitably scaled. The method makes use of a Model II analysis of variance, with certain modifications, and provides statistical tests which are not available for the more general methods of factor analysis.

I. PROBLEM

The close analogy between factor analysis and certain forms of analysis of variance has been pointed out by Burt [5, 6], and elaborated by Creasy [7]. The conditions necessary for the relations between the two to hold good are that the tests be classifiable *a priori* in a factorial or hierarchical design and that the persons (i.e., subjects) be treated as providing a mode of classification in the analysis of variance. The special role played by the hierarchical classification of tests in connection with factor analysis has long been recognized by Burt. He noted that the factor matrix obtained by means of a principal component analysis from a set of tests classifiable hierarchically can generally be rearranged to reveal a dichotomous sign pattern¹, such as that given by the transformation matrix set out below (*P*, page 155). His explanation of this pattern in terms of factors arising from the dichotomous classification of tests at each level in the hierarchy is in essence equivalent to the explanation in terms of components of random variation presented in the following discussion.

Since the individuals tested must be considered as forming a *sample*, the use of analysis of variance for this purpose is a special case of what Eisenhart [8] has termed a 'Model II Analysis', i.e., an analysis into components of variance. From the implications of such a model it will be shown in the present paper that the connection between analysis of variance and factor analysis may be not merely an analogy, but a rigorous formal relation. But the connection becomes evident only when a clear distinction is drawn between factor analysis as a 'structural' analysis and factor analysis as a form of 'discriminal' analysis.

By 'structural' analysis is meant a method which attempts to make casual statements about test performances by assigning to definite sources the covariation which arises between certain psychological tests : this was the original use

¹ See Tables I and III in the article on 'The Sign Pattern of Factor Matrices', this *Journal*, VII, pp. 15 f., and references *ibid*, p. 29.

of factor analysis. In its subsequent application to the construction of test batteries factor analysis was also used to assess whether tests of known measurement error yield reliable distinctions between individuals, and, if so, in how many dimensions : it seems appropriate to designate this 'discriminal' analysis. Unlike the approach adopted in the present paper, factor analysis as ordinarily applied does not separate these two uses, nor does it give clear-cut answers in either. The difficulty is partly statistical, particularly in multiple factor analysis where the indefiniteness of the procedure (the communality problem) leads to intractable sampling problems ; but it is also psychological, in that the interpretation of the factors depends heavily upon the composition of the original test battery, which in many cases is highly arbitrary.

It is of considerable interest, therefore, that by choosing tests in the light of specific hypotheses about their composition, and suitably adjusting their psychometric characteristics, we may make use of a Model II analysis of variance to examine the structural and discriminial properties of tests in a way free from difficult statistical or interpretational problems. Perhaps the most interesting case of such an analysis is one in which the tests are arranged in a dichotomous factorial classification according to their objective features. For example, eight classes of arithmetic tests could be constructed from the following two-fold classifications :

- A. Operation : (1) Multiplication, (2) Division.
- B. Form of problem : (1) Tabular (calculation), (2) Statement (reasoning).
- C. Required response : (1) Multiple choice, (2) Free response.

Factorial classification is of special interest in test development because it arises naturally when a test battery is constructed by systematically contrasting various features in a domain of tasks (see numerical example below). For the analysis of the factorial design it must be assumed that the variances of the test scores are equal in the population of individuals. It is also convenient, though not essential, for the scores to be adjusted to zero mean in the population. This can be arranged by using tests with standardized scores. For the discriminial part of the analysis we must also assume that the lengths of the tests have been manipulated so that the variances for the replication error (measurement error) of the tests are equal in the population of replications. Methods for this purpose based on the Spearman-Brown prophecy formula are well known in the field of test construction [10]. For sake of definiteness, replication may be defined as the administration of parallel forms of the test to the same individuals on two or more occasions. In this case the measurement error will include variance due to the (stratified) sampling of items in the construction of the parallel forms. Tests with equal replication error variance may be called 'stabilized'. Psychological tests are perhaps unique among sources of statistical data in that the investigator has freedom to adjust the origin, unit, and sometimes even the shape of the population distribution.

II. THE MODEL

To facilitate discussion, consider a design of $2^2 = 4$ tests from two dichotomies, say B and C. The tests may be identified by the ordered pairs ;

		B	
		1	2
C	1	11	12
	2	21	22

This example will show most of the important features of the 2^n design.

Assuming standardized, stabilized scores, represent the model for both the structural and discriminial analysis by the observational equation

$$x_{ijkt} = \alpha_i + \beta_{ij} + \gamma_{ik} + \delta_{ijk} + \epsilon_{ijkt},$$

where x_{ijkt} is the score of individual i on test jk on occasion t , α_i is a component of score specific to individual i and general to all tests, β_{ij} , γ_{ik} are components of score specific to the performance of individual i on tests of the B and C classification respectively, δ_{ijk} is a component of score for individual i specific to each test, and ϵ_{ijkt} is a replication error specific to each individual, test, and occasion. Other sources of variation, such as individual by occasion interaction, are believed to be of lesser importance and are omitted from the model. In some cases, however, it may be necessary to remove systematic occasion effects. Since the tests are standardized, fixed components attributable to the general difficulty of the tests and to interaction between tests do not appear. If the absolute levels of the tests are of interest, the fixed components may easily be included.

The distributions of the random components are assumed independent and normal :

$$\begin{array}{lll} \alpha_i : & N(0, \sigma_a^2) & i = 1, 2, \dots N \\ \beta_{ij} : & N(0, \sigma_b^2) & j = 1, 2 \\ \gamma_{ik} : & N(0, \sigma_c^2) & k = 1, 2 \\ \delta_{ijk} : & N(0, \sigma_d^2) & t = 1, 2, \dots n \\ \epsilon_{ijkt} & N(0, \sigma_e^2) & \end{array}$$

In the general model the number of random components with distribution over individuals equals the number of tests, i.e.,

$$1 \binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = (1+1)^n = 2^n.$$

There is also one random component, ϵ_{ijkt} , whose distribution is assumed to range over both individuals and replications. To emphasize the distinctness of these independent distributions we shall refer to the distribution over persons as the 'sampling' distribution, and that of ϵ_{ijkt} over replications and persons as the 'error' distribution. Because the number of components with a

distribution over persons only is equal to the number of tests in the dichotomous factorial design, the covariance structure of the tests may be fully explicated. This is not true in designs with more than two tests per classification, though an analysis of such designs by the methods proposed is still informative.

The expectation of the observations over replications is

$$E(x_{ijkl}) = \alpha_i + \beta_{ij} + \gamma_{ik} + \delta_{ijk}.$$

For N individuals there are thus $9N$ parameters in the model but only $4N$ observational equations. The parameters are therefore not separately estimable; but we shall see presently that four useful linear functions of the parameters for each individual are estimable. If the observations are constrained to sum to zero in the sample, only $4(N-1)$ of these functions will be estimable, but this difficulty can be avoided by estimating differences between individuals, only $N-1$ of which are independent.

III. STRUCTURAL ANALYSIS

The purpose of the structural analysis is to test whether the sample covariances between the tests are in accordance with the model. For the sake of generality we may assume that the covariance matrix for a sample of N individuals is computed from the means of r replicate scores. For the structural analysis it suffices to consider only one replication; but the more general formulation is useful for the subsequent discussion of discriminant analysis. The sample covariance matrix may be represented as

$$S_r = \frac{X'X - M_r}{N-1},$$

where X is the $N \times 4$ matrix of the means of r replicate scores and M_r is the matrix of corrections to the sample means. In the appendix it is shown that the expected covariance matrix in the population of individuals is

$$E_i(S_r) = \sum_r = \begin{bmatrix} v & a & b & c \\ a & v & c & b \\ b & c & v & a \\ c & b & a & v \end{bmatrix}$$

Here $v = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_d^2 + \sigma_e^2/r$; $a = \sigma_a^2 + \sigma_b^2$; $b = \sigma_a^2 + \sigma_c^2$; $c = \sigma_a^2$.

This form of covariance matrix Bargmann calls the 'equipredictability pattern', since it provides a necessary and sufficient condition for the equality of the multiple correlations between each variable and all others [3]. It may be

verified directly that a matrix of this form, under pre-and post-multiplication by the orthogonal matrix,

$$P = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

will reduce to its canonical (or diagonal) form, viz.,

$$P\Sigma_r P' =$$

$$\begin{bmatrix} 4\sigma_a^2 + 2\sigma_b^2 + 2\sigma_c^2 + \sigma_d^2 + \sigma_e^2/r & & & \\ & 2\sigma_b^2 + \sigma_d^2 + \sigma_e^2/r & & \\ & & 2\sigma_c^2 + \sigma_d^2 + \sigma_e^2/r & \\ & & & \sigma_d^2 + \sigma_e^2/r \end{bmatrix}$$

P does not diagonalize the *sample* covariance matrix : only the unique matrix of latent vectors computed from the sample matrix has this property. However, if the model holds, the off-diagonal elements remaining after transformation of the sample covariance matrix by P will be attributed simply to variation due to the sampling of individuals. Thus a statistical test of the hypothesis that the off-diagonal elements of the transformed sample matrix are zero in the population will constitute a test of the correctness of the model. For this purpose a likelihood ratio test is given by Wilk's Criterion ; and a χ^2 approximation for moderate to large samples has been provided by Bartlett [2] :

$$\chi^2 = -\left(N - \frac{2p+11}{6}\right) \log_e |R_r|; \quad \text{d.f.} = p(p-1)/2$$

where p is the number of variates and R_r the correlation matrix corresponding to the sample covariance matrix *after* the latter is pre- and post-multiplied by P . The determinant $|R_r|$ is most easily computed by reducing R to triangular form by the forward solution of the Gauss-Doolittle method for solving simultaneous linear equations, and multiplying together the diagonal elements of the triangular form (see [14], p. 30). Alternatively the calculation of R_r can be avoided by using the identity

$$|R_r| = |S_r| / |D_r|,$$

where

$$|D_r| = (p_1 S_r p_1') (p_2 S_r p_2') (p_3 S_r p_3') (p_4 S_r p_4').$$

The vectors p_1, p_2, p_3 , and p_4 are the rows of P . Since $|S_r| = |P\hat{S}_r P'|$, the determinant in the numerator can be computed either from the sample covariance matrix or its transform.

If the criterion of the structural test does *not* reach significance, we may accept the assumption that the composition of the scores specified by the model is correct ; and the covariation of any pair of tests can be explained in terms of

the components of the model which they share. Common components associated with ways of classifying the tests can be named and interpreted in terms of the contrasts in the factorial design. The general component cannot of course be interpreted in terms internal to the design; but in a larger set of tests it may appear as a contrast in some higher order of classification, and thus becomes interpretable without reference to external criteria.

If the criterion *does* reach significance, we take this as evidence either that the parameters cannot be assumed to be uncorrelated or that common features appear in the tests other than those specified by the classification. As an example of the latter, tests 11 and 22 should share, according to the model, only the component of score general to all tests. In reality there may be an unsuspected similarity between these tests which contributes to their co-variation, but not to that of 12 and 21. Then element 1, 4 of the covariance matrix would be significantly greater than 2, 3; and the criterion would reject the null hypothesis. In this case the factorial classification would not provide a complete characterization of the covariance structure.

Up to this point the analysis has been merely qualitative, and gives only an identification of the sources of covariation. For a quantitative statement of the importance of these sources we must estimate the components of variance σ_a^2 , σ_b^2 , σ_c^2 , and σ_d^2 . Here the connection between the structural analysis and the Model II analysis of variance becomes evident. Taking the individuals as a random dimension, we may show by expanding the usual formulae of analysis of variance that the mean squares for the random effects are simply r times the diagonal elements of PS, P' (see Table I). Thus the existing statistical theory for Model II is available for estimating and testing the significance of the required components of variance. The necessary calculations comprise the discriminial analysis, and are discussed in the next section.

Similarly, the mean squares for the fixed test effects, including the over-all mean (correction term), are found to be r times the transform of the correction matrix by P , i.e., rPM, P' . The fixed effects will be non-significant if the test scores are adjusted to zero in the population; otherwise they may be significant, and may be interpreted in terms of test difficulties or their interactions. Since estimation and testing of fixed effects are part of the classical analysis of variance, they will not be considered further in this paper.

IV. DISCRIMINAL ANALYSIS

The following discussion makes use of the vector representation of test scores which has become familiar in both factor analysis and multivariate analysis. In the case of the 2^2 test design, the four standardized and stabilized scores for each individual can be regarded as the coordinates of a vector in the 4-space of tests. We assume that these vectors are distributed according to the multivariate normal distribution

$$x_i : N(O, \Sigma_r) \quad .$$

$\begin{matrix} 1 \times 4 & 4 \times 4 \end{matrix}$

In this distribution the contours of uniform density are, in general, families of ellipsoids centring about O . The orthogonal transformation P refers these ellipsoids to their principal axes, and puts the distribution function in its canonical form, viz.,

$$x_i P' : N \left(\begin{matrix} O \\ 1 \times 4 \end{matrix}, \begin{matrix} P \Sigma_r P' \\ 4 \times 4 \end{matrix} \right) .$$

The canonical form of the covariance matrix is, as we have seen, a diagonal matrix whose elements are proportional to the squared distance in the direction of the principal axes from the origin to the surface of a given ellipsoid. It is well known that these distances, which are proportional to latent roots of the matrix Σ_r , include the greatest and least distance possible in the space of the distribution [12].

From the expected mean squares of $P \Sigma_r P'$, we observe that under the model variation in the space of the tests is made up of two parts, one spherical and proportional to σ_e^2 in all directions, the other in general ellipsoidal and proportional to $4\sigma_a^2 + 2\sigma_b^2 + 2\sigma_c^2 + \sigma_d^2$, $2\sigma_b^2 + \sigma_d^2$, $2\sigma_c^2 + \sigma_d^2$ and σ_d^2 in the (orthogonal) directions of the respective principal axes. For the purpose of discriminating among individuals the spherical part is of no interest since it represents random error from occasion to occasion. It measures the extent to which we *cannot* claim that the test scores provide a quantitative characterization of the individuals persisting over the interval of replication. Just how long a replication interval is indicated will depend of course on the purpose of the research. To estimate this error we may use replications obtained from the individuals under investigation, and compute v_e , possibly adjusting for systematic occasion effects, as indicated in Table I. In many cases, however, squared standard errors of measurement derived from existing parallel-form reliabilities will serve. Using

TABLE I. MODEL II
ANALYSIS OF VARIANCE FOR MEANS OF r REPLICATIONS OF FOUR FACTORIALLY CLASSIFIED TESTS

Source of Variation	d.f.*	Mean Squares	Expected Mean Squares
Individuals			
A	$N-1$	$v_a = p_1 S_r p_1'$	$\sigma_e^2/r + \sigma_a^2 + 2\sigma_c^2 + 2\sigma_b^2 + 4\sigma_d^2$
B	$N-1$	$v_b = p_2 S_r p_2'$	$\sigma_e^2/r + \sigma_a^2 + 2\sigma_b^2$
C	$N-1$	$v_c = p_3 S_r p_3'$	$\sigma_e^2/r + \sigma_a^2 + 2\sigma_c^2$
D	$N-1$	$v_d = p_4 S_r p_4'$	$\sigma_e^2/r + \sigma_d^2$
Replication			
Error**	$4N(r-1)$	$v_e = \frac{1}{r} \sum_{i,j,k} (\sum_t x_{ijkt}^2 - r x_{ijk}^2)$	σ_e^2/r
Total	$4Nr-1$	$\sum_{i,j,k,t} x_{ijkt}^2 - 4Nr \bar{x}^2$	

* Assuming S_r corrected for the sample means.

** It may be necessary to remove over-all occasion effects, in which case the d.f. become $4(N-1)(r-1)$.

reliabilities is equivalent to removing systematic occasion effects ; and the degrees of freedom of the corresponding error estimate is equal to $4(N-1)$. For this purpose Kuder-Richardson and split-half reliability are unsuitable.

Tests of significance for the non-error components are available [1]. For the B, C, and D components these are simple F tests as indicated in Table II. For the general component, however, there is no appropriate denominator for an exact F test. A useful approximate test is available based on results due to Satterthwaite, which show that a ratio derived from a linear compound of mean squares as a denominator is distributed in good approximation to F when the effective degrees of freedom for the denominator are properly calculated [15]. The test for the 2^2 design is shown in Table II ; the general approach may be found in ([1], p. 350).

An important consideration in the analysis is the psychological interpretation of the different components of the variance. In this connection we note that the transformation of the covariance matrix by P is equivalent to transforming the original scores to a set of scores uncorrelated in the population. Let the transformed scores for individual i be

TABLE II. VALUES FOR F-RATIOS AND OTHER STATISTICS
Effect

	A	B	C	D
F -ratios	$\frac{v_a}{v_b + v_c - v_d}$	$\frac{v_b}{v_d}$	$\frac{v_c}{v_d}$	$\frac{v_d}{v_e}$
Degrees of Freedom (n_1)	$N-1$	$N-1$	$N-1$	$N-1$
Degrees of Freedom (n_2)*	$\frac{(v_b + v_c - v_d)^2}{(v_b^2 + v_c^2 + v_d^2)/(N-1)}$	$N-1$	$N-1$	$4N(r-1)$ or $4(N-1)(r-1)$
Variance Components	$\hat{\sigma}_a^2 = \frac{v_a - v_b - v_c + v_d}{4}$	$\hat{\sigma}_b^2 = \frac{v_b - v_d}{2}$	$\hat{\sigma}_c^2 = \frac{v_c - v_d}{2}$	$\hat{\sigma}_d^2 = v_d - v_e$
Intra-class Correlations	$\frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_b^2 + \hat{\sigma}_c^2 + \hat{\sigma}_d^2 + \hat{\sigma}_e^2/r}$	$\frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_d^2 + \hat{\sigma}_e^2/r}$	$\frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \hat{\sigma}_d^2 + \hat{\sigma}_e^2/r}$	$\frac{\hat{\sigma}_d^2}{\hat{\sigma}_d^2 + \hat{\sigma}_e^2/r}$
Least Significant	$t_{.05} \sqrt{(v_b + v_c - v_d)}$	$t_{.05} \sqrt{(v_d)}$	$t_{.05} \sqrt{(v_d)}$	$t_{.05} \sqrt{(v_e)}$

* The expression (in col. A) is based on Satterthwaite's approximation [15].

$$u_{i1} = p_1 x_i',$$

$$u_{i2} = p_2 x_i',$$

$$u_{i3} = p_3 x_i',$$

$$u_{i4} = p_4 x_i'.$$

These scores estimate various linear combinations of the parameters in the model, as indicated by their expectation over replications,

$$E(u_{i1}) = 2a_i + \beta_{i1} + \beta_{i2} + \gamma_{i1} + \gamma_{i2} + \frac{1}{2}(\delta_{i11} + \delta_{i12} + \delta_{i21} + \delta_{i22}),$$

$$E(u_{i2}) = \beta_{i1} - \beta_{i2} + \frac{1}{2}(\delta_{i11} + \delta_{i12} - \delta_{i21} - \delta_{i22}),$$

$$E(u_{i3}) = \gamma_{i1} - \gamma_{i2} + \frac{1}{2}(\delta_{i11} - \delta_{i12} + \delta_{i21} - \delta_{i22}),$$

$$E(u_{i4}) = \frac{1}{2}(\delta_{i11} - \delta_{i12} - \delta_{i21} + \delta_{i22}).$$

The variances of these scores, calculated over the individuals, form the respective mean squares shown in Table I.

In the absence of effects due to the test classification and specificity (i.e., if $\hat{\sigma}_b^2$, $\hat{\sigma}_c^2$, and $\hat{\sigma}_d^2$ are not significant), u_{i1} estimates separately the level of general ability of individual i with respect to these tests (or, strictly speaking, $u_{i1} - u_{ij}$ estimates $\alpha_i - \alpha_j$ when the x_i are corrected to zero in the sample). If effects other than the general are significant, u_{i1} can be considered to estimate α_i only in a weak sense: that is, a difference between the scores of two randomly selected individuals can be interpreted as a difference in general ability only if it is significantly in excess of differences which might be expected from the remaining sources of independent random variation. Frequently in mental testing, the variation of general performance among persons is so much larger than the variation due to test effects that the mean for all tests, while it may err for particular individuals, is on the whole a good characterization of general ability.

In the absence of specific effects, u_{i2} and u_{i3} estimate separately contrasts in abilities defined by the B and C dichotomies of the test classification. The signs of these scores indicate on which category of the dichotomies the individual excels. Again, this interpretation is attenuated by significant $\hat{\sigma}_d^2$, but if the specific effect is relatively small, the estimates of $\beta_{1i} - \beta_{2i}$ and $\gamma_{1i} - \gamma_{2i}$ may on the whole be useful.

Finally, u_{i4} estimates separately a function of the specific effects. Its interpretation is clarified if we write its expectation as a difference of differences,

$$\frac{1}{2} (\delta_{i11} - \delta_{i12}) - \frac{1}{2} (\delta_{i21} - \delta_{i22}).$$

A positive u_{i4} indicates that specific effects are causing test scores in the first categories of the B classification to differ more than those in the second. Replication error also contributes specific variation to the tests, but unlike the D effect, it can be suppressed by increasing the number of replications. Note that, if there were no significant effects other than the specific, all the transformed scores would significantly discriminate among the individuals relative to replication error. But in this case transforming the scores would add nothing to the interpretation, and we might just as well report the original scores.

The usefulness of the transformed scores in making interpretable discriminations may be expressed in several ways, as indicated in Table II. Variance components can be estimated from the sample mean squares. Such estimates have been shown to be unbiased and of minimum variance [9]. As an aid in comparing the components we may compute the generalized intraclass correlations, giving the ratio of the effect variance to the sum of the variances from all sources [11]. An intuitively more direct index, however, is provided by the least difference in score between two individuals which can be attributed to the effect. It is obtained by inverting the hypothesis of no significant effect assuming $N=2$ and may be expressed in terms of the t distribution, with degrees of freedom equal to that of the denominator of the corresponding F -test. If this difference is small relative to the total range of differences among individuals, the score may be considered useful.

The foregoing discussion assumes that the hypothetical structure of the population covariance matrix has not been contradicted by the structural analysis of the sample matrix. If it has, the discriminant analysis can still be carried out, but its interpretation is limited. Significance of all the random components is still a necessary condition for reliable distinctions in all possible directions in the space of tests, but it is no longer sufficient. If the P transformation cannot be assumed to refer the population covariance matrix to principal axes, there still exists some (unknown) transformation which does. This transformation would in general result in more extreme maximal and minimal mean squares, and the latter may then prove insignificant. Also, if the model does not hold, the interpretation of the transformed scores becomes uncertain.

V. NUMERICAL EXAMPLE

The sample covariance matrix and correlation matrix for four subtests of an experimental form of the Language Modalities Survey (to be used with aphasic subjects) is shown in Table III [13]. The tests are classified according as the stimulus is a word or a picture and the response is oral or graphic :

		Response	
		Oral	Graphic
Stimulus	Picture	11	12
	Word	21	22

The tests were not standardized in the population of subjects; but the responses were scored quantitatively according to degree of correctness, and so scaled that the sampling and error variances were roughly constant. The transformation of the covariance matrix by P is shown in Table III along with the corresponding correlation matrix. It will be seen that the off-diagonal elements in the correlation matrix becomes small and more or less centered on zero. For the structural test the determinant of S , was computed (using the Cholesky, or 'square root', factorization rather than the slower standard elimination method [4]).

The approximation for the likelihood ratio test was calculated as shown in Table III. The hypothesis of zero association was not rejected, suggesting that the discriminant analysis should proceed. The analysis of variance which

TABLE III. STRUCTURAL ANALYSIS

1. *Covariance and correlation matrices for the original mean scores ($r=2$)*

		Covariances (S_r)				Correlations			
Tests	11	12	21	22	11	12	21	22	
11	7.251	3.113	4.844	3.133	1.000	.315	.524	.314	
12		13.483	5.972	9.245		1.000	.473	.681	
21			11.802	3.209			1.000	.252	
22				13.687				1.000	

2. *Transformed covariance and correlation matrix*

Tests	Covariances (PS_rP')				Correlations			
	11	12	21	22	11	12	21	22
11	26.313	-1.237	-4.230	-2.506	1.000	-.137	-.250	-.201
12		3.120	.333	.172		1.000	.057	.040
21			10.887	-1.141			1.000	-.142
22				5.903				1.000

3. *Tests of association in the transformed covariance matrix*

$$\chi^2 = \frac{1}{2} p(p-1) - (N - \frac{2p+11}{6}) \log_e \left| \frac{S_r}{D_r} \right|$$

$$\chi^2 = \frac{1}{2} 4(4-1) - (50 - \frac{2 \times 4 + 11}{6}) (2.302585) \log_{10} \frac{(7.2508) (13.4828) (11.8024) (13.687)}{(26.313) (3.120) (10.887) (5.903)}$$

$$\chi = 7.853 \quad p \approx 0.25$$

resulted is shown in Table IV. The F tests indicate significant individual differences in the general (A), response (C), and specific (D) effects, but none in the stimulus (B) effect. Thus the transformed scores u_1^i and u_3^i subsume information from common effects in the tests, while u_2^i and u_4^i are linear functions only of specifics. The substantive interpretation of the transformed scores is that the aphasic subjects differ individually in general ability in the four tests, in oral as opposed to graphic abilities for both word and picture stimuli, and in their pattern of performance on the separate tests, but not in their ability to respond to picture as opposed to word stimuli. The substantial specific variance implies that more is involved in the responses to these tests than mere stimulus-response modality effects. Other evidence suggests that a symbol transformation effect also might be postulated.

TABLE IV. ANALYSIS OF VARIANCE ($N=50$, $r=2$)

Effect	d.f.	Mean Squares	F	Variance Components	Intraclass Correlations	Least Significant Differences
A	49	26.313	3.25*	4.552	.351	5.95
B	49	3.120	.53	—	—	—
C	49	10.887	1.844**	2.492	.297	4.88
D	49	5.903	3.36***	4.145	.702	2.61
Error	196	1.758				

* Effective d.f. of denominator = 20 : significant.

** Significant at 0.05 level.

*** Significant at 0.01 level.

VI. OTHER MODELS

The analysis outlined in this paper is a particular result of the general theory of linear estimation and tests of hypotheses. The theory can be extended to other designs, which are plausible for psychological tests, and may be used in a similar way to deduce the expected form of the covariance matrix or to devise structural and discriminial tests. A number of such designs are being explored, and a monograph reporting the results and the steps in the general method is in preparation.

APPENDIX : EXPECTATION OF THE SAMPLE COVARIANCE MATRIX

To ease the burden of notation the following computation is limited to the case already discussed. Write the model for means of r replications as follows :

$$x_{ijk} = \alpha_i + \beta_{ij} + \gamma_{ik} + \delta_{ijk} + \epsilon_{ijk},$$

and make the same distribution assumptions as before. The observational equations can then be expressed in matrix notation by defining (i) the model matrix as

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and (ii) the $N \times 13$ matrix of parameters and errors as

$$Z = [\alpha_i, \beta_{i1}, \beta_{i2}, \gamma_{i1}, \gamma_{i2}, \delta_{i11}, \delta_{i12}, \delta_{i21}, \delta_{i22}, \epsilon_{i11}, \epsilon_{i12}, \epsilon_{i21}, \epsilon_{i22}]$$

where $i = 1, 2, \dots, N$.

Then the observations may be denoted by $X' = AZ'$, where A and Z' are 4×13 and $13 \times N$ respectively, and the expected sum of squares and cross products by

$$E_i[X'X] = A[E_i(Z'Z)]A'.$$

Calculating the various expectations in $Z'Z$, we find $E_i(Z'Z)$ is a 13×13 diagonal matrix with terms

$$N[\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2, \sigma_f^2, \sigma_g^2, \sigma_h^2, \sigma_i^2/r, \sigma_j^2/r, \sigma_k^2/r, \sigma_l^2/r].$$

The expectation of the correction matrix, $E_i(M_r) = AE_i \frac{(Z'Z)}{N} A'$, yields a similar matrix without the multiplier N . Hence,

$$\begin{aligned} E_i(S_r) &= E_i(X'X - M_r)/(N-1) \\ &= A[(E_i(Z'Z) - E_i \frac{(Z'Z)}{N})] A'/(N-1), \end{aligned}$$

that is, pre- and post-multiplication of the expectations by A gives the expected value for the sample covariance matrix exhibited in the body of the paper. The same result applies when the fixed parameters are included in the model. The sums of squares for the fixed effects vanish on subtracting the correction matrix.

REFERENCES

- [1] ANDERSON, R. L. and BANCROFT, T. A. (1952). *Statistical Theory in Research*. New York : McGraw-Hill Book Company.
- [2] ANDERSON, T. W. (1958). *An Introduction to Multivariate Analysis*. New York : Wiley.
- [3] BARGMANN, ROLF. (1957.) *A Study of Independence and Dependence in Multivariate Normal Analysis*. Mimeograph Series No. 186, Institute of Statistics, University of North Carolina.
- [4] BODEWIG, E. (1956). *Matrix Calculus*. New York : Interscience.
- [5] BURT, CYRIL (1940). *The Factors of the Mind*. London : University of London Press.
- [6] BURT, CYRIL (1947). Factor analysis and analysis of variance. *Brit. J. Psychol., Stat. Sect.*, I, 3-26.

- [7] CREASY, MONICA A. (1957). Analysis of variance as an alternative to factor analysis. *J. Roy. Stat. Soc., Series B*, XIX, 318-325.
- [8] EISENHART, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, IV, 1-21.
- [9] GRAYBILL, F. A. and WORTHAM, A. W. (1956). A note on uniformly best unbiased estimators for variance components. *J. Amer. Stat. Assoc.*, LI, 266-268.
- [10] GULLIKSEN, H. (1950). *Theory of mental tests*. New York : Wiley.
- [11] HAGGARD, E. A. (1958). *Intraclass correlation and the analysis of variance*. New York : Dryden Press.
- [12] HOTELLING, HAROLD (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, XXIV, 360-378.
- [13] *Language Modalities Survey : A Test for Aphasia*. (1959.) Speech Clinic, University of Chicago and Psychometric Laboratory, University of North Carolina, Chapel Hill.
- [14] RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York : Wiley.
- [15] SATTERTHWAITE, F. E. (1941). Synthesis of variance. *Psychometrika*, VI, 309-316.

DISCREPANCIES IN THE VARIANCES OF TEST RESULTS FOR NORMAL AND NEUROTIC CHILDREN

By A. E. MAXWELL

Institute of Psychiatry, Maudsley Hospital, University of London

When the scores obtained with the WISC battery of tests from children attending a psychiatric clinic were compared with those obtained from normal children, they showed a greater dispersion. Attempts to explain the discrepancy in terms of selection, non-normality, or emotional and environmental factors, proved unconvincing. An application of Bartlett's formulae for the sampling theory of ability suggested that the difference might be due to genetic factors arising from pronounced assortative mating.

I. PROBLEM

Elsewhere [8] I have reported discrepancies between the patterns of abilities for normal children and for children attending a psychiatric clinic respectively. It was noted that for the latter, the majority of whom were referred for reasons other than abnormal cognitive development, the dispersions of scores on WISC subtests [15] were greater than the expected values. In addition the inter-correlations between the tests were greater than those given by Wechsler [15]; and the results were reflected in a factor analysis in which the general factor was considerably enhanced. Although attempts were made to explain the discrepancies in dispersion in terms of selective sampling, heterogeneous groups within the psychiatric sample, non-normal distribution of test scores, and such environmental and emotional factors as can arise from unsatisfactory parental and social relations or disturbances of function, no satisfactory explanation was found. In this paper further evidence of the excessive dispersion with the WISC subtests is given; and a possible genetic explanation will be suggested.

II. THE DATA

Whereas in the earlier study the age of the psychiatric children ranged from 8 to 13 years, in this study attention was confined to children between $9\frac{6}{12}$ and $10\frac{5}{12}$ years. This range was chosen because control data for 80 normal children of similar age were available from work carried out by Dr. Jones at University College [5]. From the files of the Maudsley Hospital Children's Department we extracted data for children within this age range for whom complete records for the 10 subtests of the WISC here considered were available. This yielded a sample of 59 boys and 36 girls.

In the first instance means and standard deviations were calculated separately for boys and girls; but, apart from a greater mean on Block Design for boys

and a tendency for the standard deviations for girls to exceed those for boys, the results were similar. Both raw and scaled scores were tabulated; and the means and standard deviations for the combined samples are given in Table I. With the scaled scores all the means (except Vocabulary) are below the expected values of 10; and all (except Similarities) are significantly lower than those obtained by Dr. Jones. All the standard deviations for scaled scores (except Digit Span) are greater than 3—significantly so when tested by chi squared. In addition the variances of four of the tests, Comprehension, Similarities, Vocabulary and Block Design, differ significantly from those reported by Jones; and all our standard deviations are numerically greater than hers. Provided the distributions are normal this suggests that the population from which our sample is drawn has greater dispersion of test scores than the normal population.

Now Jones, when choosing her children, aimed at a stratified sample in which the proportion of children from each socio-economic group was as near as possible to that obtaining in the general population. From the schools available she had difficulty in getting the necessary quota for the lower income groups; and the relatively high means (Table I) which she obtained may reflect this difficulty. Possibly for the same reason her standard deviations tended to be lower than the expected values; but she herself attributed this in part to the homogeneous educational background of the children. However, the same homogeneous background can be claimed for those in our psychiatric sample. Moreover, it is known that for the hospital population from which it is drawn the proportion of children from the two highest socio-economic groups is about 21·4 per cent as compared with 17·4 per cent for the general population, so that our sample seems to compare well with hers as regards type of home. Since the transformation of raw to scaled scores should have equated the two samples as regards the dispersions for the subtests, we can conclude that the population from which our sample is drawn has a greater dispersion than hers.

TABLE I. MEANS AND STANDARD DEVIATIONS FOR 10 WISC SUBTESTS

<i>Subtests</i>	Psychiatric Sample		Psychiatric Sample		Normal Sample	
	<i>Raw Scores</i>		<i>Scaled Scores</i>		<i>Scaled Scores</i>	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Comprehension	9·95	3·70	8·62	3·34	10·1	2·6
Arithmetic	7·41	2·53	8·30	3·65	11·5	3·0
Similarities	8·95	3·94	9·94	3·62	9·9	2·3
Vocabulary	30·61	9·80	10·05	3·78	11·0	2·6
Digit Span	8·13	2·13	8·31	2·86	10·3	2·8
Picture Completion	10·35	3·11	9·94	3·63	11·0	2·9
Picture Arrangement	21·39	9·55	8·47	3·28	9·5	2·8
Block Design	17·80	10·91	9·53	3·62	11·7	2·8
Object Assembly	18·37	5·80	9·07	3·42	10·3	2·9
Coding	26·24	11·05	7·67	3·41	10·3	2·8

Furthermore, lest it should be argued that the differences in dispersion between the two samples are due to bimodality in the frequency distribution, or other non-normal effect, a normal curve was fitted to the distribution of the

III. CORRELATIONS AND FACTORS

TABLE II. CORRELATION COEFFICIENTS

[illegible]

With the advent of electronic computers it is now feasible to use factor analytic methods which give statistically efficient estimates of factor loadings, despite the heavy calculations involved. Lawley's Maximum Likelihood Method, for which a programme [9] is now available, was employed. The results for the raw scores are given in Table III, together with the results of a similar analysis of Jones' data.

With our data three significant factors were found. Lawley's χ^2 test of the residuals gave a value of 44.8, based on 26 degrees of freedom, after two factors had been fitted, and one of 15.9, based on 18 degrees of freedom, after three factors had been fitted. The former is significant beyond the 2.5 per cent level; the latter is not significant. By contrast only two significant factors were found with Jones' data.

TABLE III. UNROTATED FACTOR LOADINGS

Subtest	(a) Psychiatric Sample				(b) Normal Sample		
	i	ii	iii	h^2	i	ii	h^2
i	.782	.289	-.003	.695	.658	.279	.511
ii	.773	-.046	-.267	.671	.645	-.090	.424
iii	.865	.198	.069	.792	.763	-.056	.585
iv	.883	.242	.130	.855	.849	.285	.802
v	.637	-.061	-.395	.566	.512	-.125	.278
vi	.771	-.106	-.050	.608	.588	.022	.346
vii	.719	-.063	-.119	.535	.653	.001	.426
viii	.738	-.466	.170	.791	.708	-.323	.606
ix	.726	-.311	.158	.649	.598	-.371	.496
x	.678	-.152	-.319	.585	.464	-.410	.383
Variance (Percentage)	57.86	5.39	4.21	67.47	42.60	5.97	48.57

Inspection of Table III shows that the general pattern of loadings is similar in both studies. For the first factor the highest loadings are those for Vocabulary, Similarities, Comprehension and Arithmetic, and for the second those for Block Design and Object Assembly. The fact that for our sample the first factor has a larger variance and a third factor emerges results from the increased size of the correlations; and they in turn result from the greater dispersion of the scores. Elsewhere [8] similar discrepancies between results for psychiatric and normal samples are reported; but efforts to explain them in terms of selection, or other familiar notions were not convincing. Owing to the parallelism between the two studies such explanations are not acceptable here either; and other explanations must be sought.

IV. THE SAMPLING THEORY

Thomson depicted the human mind, as a complex of neural activity involving a large number, N , of elements "identified, on the mental side, with something of the nature of Thorndike's 'bonds', and, on the bodily side with neurone arcs". (These we may assume are, in part at least, genetically determined.)

He then postulated that the performance of any given test 'sampled' a fixed proportion p of these components. If the effects of the components were taken to be equal, independent, and additive, and if two tests sampled proportions p_1 and p_2 , then the expected value of the correlations between the tests would be given by the formula

$$r_{12} = \frac{p_1 p_2 N}{\sqrt{(p_1 N p_2 N)}} = \sqrt{(p_1 p_2)}. \quad (1)$$

Now if a number of tests ($i=1, 2, \dots, k$) are concerned, sampling proportions $p_1, p_2, \dots, p_k$, respectively of the elemental components, and their intercorrelations are obtained by eqn. (1), the resulting matrix of correlations would be hierarchical in Spearman's sense, and would lead to one general and k specific factors. However, if some of the tests, say 1 and 2, sampled only a subpool or fraction p of all the N components, their intercorrelations would thereby be increased, and would be given by the formula

$$r_{12} = \sqrt{(p_1 p_2)/p}. \quad (2)$$

In this way group factors would arise. Formula (2) was first given by Bartlett [1] who demonstrated that, if Thomson's model were assumed, a rigorous definition of 'general', 'group' and 'specific' factors could be given in terms of it. The existence of such factors had been fairly satisfactorily demonstrated; and this adds to the plausibility of Thomson's model. Burt's hypothesis that 'intelligence' is under the control of a very large number of genes [3, 4] seems to be acceptable from the standpoint of modern genetics.

Now let us see what deductions can be made from this theory. One, due to Bartlett [2], arose as a result of an article by Price [11] in which he criticized Spearman's postulation of a general factor of human ability, g , deduced from the intercorrelations of cognitive tests. Price showed, by applying Fisher's formulae as modified by Sewall Wright, that "intercorrelations of otherwise uncorrelated traits arise as a result of cross-marital correlation", and cross-marital correlation is a consequence of assortative mating. He concluded that "the intercorrelations which g has been presumed to illumine are seen primarily as consequences of social and therefore marital importance which has attached to the abilities concerned".

Bartlett [2] also using Sewall Wright's formulae, took up Price's challenge. He agreed that "it is hardly possible to deny his conclusion that, since an individual's traits will in part be dependent upon an underlying genetic make-up, assortative mating must influence the system of correlations that may hold among them". He then extended the statistical theory of factor analysis based on Thomson's model to take account of the correlational effects demonstrated by Price. In particular he showed that, if as a result of assortative mating or some similar effect, an average correlation ρ arose between the components in Thomson's model the correlation between two tests, which sampled respectively

p_1 and p_2 of these components, would now be increased. In particular eqn. (1) would become

$$r_{12} = \sqrt{\left[\frac{p_1 \{1 + (N-1)\rho\}}{1 + (Np_1 - 1)\rho} \right]} \times \left[\sqrt{\frac{p_2 \{1 + (N-1)\rho\}}{1 + (Np_2 - 1)\rho}} \right]. \quad (3)$$

Correlations between other combinations of tests would similarly be increased; and the net effect would be one of *stabilizing* and *intensifying* the general and group factors at the expense of the specific.

V. APPLICATIONS

If we return to the factor loadings in Table III, we see that the increase in loadings observed for the psychiatric sample over those for the normal samples, and the simultaneous reduction of the specifics, is the precise effect which, Bartlett says, must result from correlation between the elemental components in the model. What we are implying is that the discrepancies between the correlational results for the two samples might be explained in terms of assortative mating (or some similar concept from the kit-bag of geneticists) on the part of the forebears of the psychiatric sample. To support this suggestion we can show, albeit indirectly, that evidence of 'assortative mating' does exist. To do this we take as an average of the factor loadings for each sample the square root of total percentage communal variance divided by 100. This gives an average factor loading of 0.821 for the psychiatric sample and one of 0.697 for the control samples.

Now Burt and Howard ([4], p. 129), have demonstrated that, on the basis of test results uncorrected for environmental effects and unreliability, about 77 per cent of the variability of children's test scores can be attributed to genetic or hereditary causes. Using this figure and the average factor loadings just given, it can be shown, by employing Bartlett's formulae [2], that for a discrepancy between the factor loadings as great as that here observed the expected correlation between the intelligence of children's parents would require to be somewhat greater than 0.70. This is much higher than the value of 0.40 to 0.50 ([3], p. 175; [10]) generally accepted as obtaining between the intelligence of parents mating randomly, so it is reasonable to deduce that something in the nature of pronounced 'assortative mating' has been at work where the parents of the children in our sample are concerned. To obtain more direct evidence it would be necessary to secure an estimate of the correlation between the intelligence of parents of children who required psychiatric help. Unfortunately there are many obstacles in the way of obtaining a reliable estimate of the former correlation. The information in the children's case histories is inadequate since as a rule only one parent accompanies the child to hospital and, a high proportion come from broken homes. Parents cannot be compelled to undergo tests of intelligence; and, though a percentage of them could be relied on to volunteer, such an estimate would be suspect. However, an unbiased estimate

of the correlation in question may be possible when parents, who themselves have taken the 11+ examination, come to hospital with their children.

If we assume that some proportion of the children in our sample are the offspring of forebears for whom assortative mating is pronounced, then these children would possess a constitution typified by genetic uniformity, and we have it on good authority that such uniformity is bad. In general such a uniformity leads to a decrease in fitness (in the geneticists' sense). The notion that the neurotic constitution is a real thing is not new; and the idea that it is "predominantly determined by a very large number of genes of small effect" has already been suggested [12]. Those who inherit this neurotic constitution, or in whom it is brought about by unwise mating, are prone to mental breakdown, or may require psychiatric help, when faced with stressful or exacting situations. It should be noted that in their paper the Slaters give evidence for greater test score dispersion for neurotic than for normal subjects.

Finally we may recall that in eqn. (1) the quantity p_1 refers to the proportion of elemental components sampled by test 1 in a hypothetical brain in which the components are assumed to be independent. But, when we came to consider a brain in which the components are correlated, the quantity p'_1 corresponding to p_1 (eqn. (3)) increased to

$$p'_1 = \frac{p_1 \{1 + (N-1)\rho\}}{1 + (Np_1 - 1)\rho}. \quad (4)$$

Since p'_1 is greater than p_1 the variance of test 1, which from eqn. (1) is seen to be $p_1 N$, will be increased. This, on our theory, furnishes an explanation of the larger standard deviations obtained by the sample of psychiatric children on the tests.

VI. SUMMARY AND CONCLUSIONS

The problem discussed in the foregoing inquiry is the unexpected increase in the standard deviations, correlations, and factor loadings obtained from the WISC intelligence tests when applied to a group of children referred for investigation to a psychiatric clinic. The view that individual differences in intelligence are largely due to the effects of numerous genes whose effects are similar, additive, and cumulative is accepted; and on this basis it is argued that the increase in the constants calculated from the test scores is attributable to pronounced *assortative mating* among parents of such children, and that this in turn would partly account for the neurotic tendencies exhibited by the children. An attempt has been made to confirm this hypothesis by applying formulae developed by Bartlett¹.

¹ I am grateful to Dr. K. Cameron for his kind co-operation in carrying out this study, to my colleagues for their valuable criticism, and to Mrs. V. Beevers for tabulating the data. The factor analysis was carried out, by permission of Dr. R. A. Buckingham and with the assistance of Mr. W. L. B. Nixon, on the London University Mercury Computer. I am indebted to Dr. Sheila Jones of University College for permission to reproduce her means, standard deviations, and correlation coefficients.

REFERENCES

- [1] BARTLETT, M. S. (1937). The statistical conception of factors. *Brit. J. Psychol.*, XXVIII, 97-104.
- [2] BARTLETT, M. S. (1937). Note on the development of correlations among genetic components of ability. *Ann. Eugen.*, VII, 299-302.
- [3] BURT, C. (1955). The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, XXV, 158-177.
- [4] BURT, C. and HOWARD, M. (1956). The multifactorial theory of intelligence and its application to intelligence. *Brit. J. stat. Psychol.*, IX, 95-131.
- [5] JONES, S. (1959). *A Statistical Study of the Wechsler Intelligence Scale for Children*. Ph.D. Thesis. University of London.
- [6] LAWLEY, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edin.*, LX, 64-76.
- [7] MAXWELL, A. E. (1959). A factor analysis of the Wechsler intelligence scale for children. *Brit. J. educ. Psychol.*, XXIX, 237-241.
- [8] MAXWELL, A. E. (1960). Discrepancies between the pattern of abilities for normal and neurotic children. *J. of ment. Science* (in press).
- [9] NIXON, W. L. B. and MAXWELL, A. E. (1960). *Mercury Programme for Lawley Maximum Likelihood Factor Analysis*. Laboratory Notes. University of London Computer Unit.
- [10] PENROSE, L. S. (1950). *Problems of Intelligence and Heredity*. Moray House, Edinburgh.
- [11] PRICE, B. (1936). Homogamy and the intercorrelation of capacity traits. *Ann. Eugen.*, VII, 22-27.
- [12] SLATER, E. and SLATER, P. (1944). A heuristic theory of neurosis. *J. of Neur., Neurosurg. & Psychiat.*, VII, 49-55.
- [13] THOMSON, G. H. (1927). A worked-out example of the possible linkages of four correlated variables on the sampling theory. *Brit. J. Psychol.*, XVIII, 68-77.
- [14] THOMSON, G. H. (1951). *The Factorial Analysis of Human Ability*. London: University of London Press (5th ed.).
- [15] WECHSLER, D. (1949). *Wechsler Intelligence Scale for Children : (Manual)*. New York: The Psychological Corporation.

DISCREPANCIES IN THE VARIANCES OF TEST RESULTS

A Note on Mr. Maxwell's Genetic Hypothesis

In view of the widespread changes that are taking place in the mating systems that are traditional in this country, any study of the effects of assortative mating, such as Mr. Maxwell has undertaken, is of special value¹. In applying his theory to explain the increased variances and correlations that he has noted among neurotic children, he has been good enough to quote our own attempts to account for somewhat similar data, obtained from normal children, and in particular our application of multifactorial theory of inheritance to variations in general intelligence. Nevertheless, there are one or two statements in his references that we should like to correct or question.

On p. 170 he quotes (from Burt's article on 'The Concept of Intelligence', *Brit. J. Educ. Psychol.*, XXV, p. 175) a value of 0.40 to 0.50 as representing the amount of correlation "between the intelligence of parents *mating at random*". This is not quite accurate. If the parents mated at random, the correlation would be zero. As will be seen from the page preceding that to which he refers, the correlation cited represents the figure for *assortative mating* (italicized) obtained in our earlier surveys. But we fully agree, as was there pointed out, that "one of the calculable results of assortative mating would be to increase the variance"; and the extent of the increase was estimated as about 22 per cent. It was, however, noted that, in the general population, as we should expect from the gradual breakdown of social and geographical barriers during recent years, the amount of assortative mating seemed to be diminishing: in later surveys the correlation between parents was distinctly lower.

In a study of neurotic children which we carried out in London schools and child guidance centres (reported in a joint paper slightly earlier than that quoted by Mr. Maxwell, *Brit. J. Statist. Psychol.*, V, pp. 39-58), we found that, if anything, parents of neurotic and maladjusted children were *less* well assorted (in intelligence and other characteristics) than parents of normal children; and this of itself appeared to be a not infrequent cause of maladjustment in the children. Mr. Maxwell, it is true, tells us (on the strength of an unnamed 'authority') that the effects on the children of marked assortative mating among parents "is bad"; but from our own experience we feel convinced that, so far as any generalization is possible, the reverse is the case.

If he is right in assuming that Dr. Sheila Jones' sample is fairly representative of the general school population, then Mr. Maxwell must surely be mistaken in concluding that his sample does not contain "an excess of bright and dull children". The distribution of Dr. Jones' sample is virtually normal, and has a standard deviation of approximately 15 S.D. Mr. Maxwell has shown by formal tests that the distribution of his own sample is also virtually normal; but the standard deviation is about 25 per cent larger. In Dr. Jones' sample therefore about 16 per cent of the children must have I.Q.'s over 115 (or under 85); but with Mr. Maxwell's sample as many as 21 per cent must have I.Q.'s over or under those figures. From this it follows that Mr. Maxwell's group might be regarded (in a slightly paradoxical sense) as a 'selected' sample drawn from the general population

¹ It should perhaps be noted that the idea of a 'sampling theory' to interpret correlations due to heredity was originally suggested by Pearson's colleague, W. F. R. Weldon ('in *Inheritance in Plants and Animals*', *ap.* T. B. Strong, *Lectures on the Method of Science*, 1906, pp. 100f.), from whom Thomson took the device of experimenting with throws of dice and whose formulae he and Brown used (see Brown and Thomson, *Essentials of Mental Measurement*, p. 138, footnote ¹).

in such a way that the standard deviation is increased rather than diminished. An application of the familiar formulae devised by Karl Pearson to 'correct' for selection seems to reduce his coefficients very nearly to the values reported by Dr. Jones.

From what is already known of the characteristics of neurotic children, it would appear that the increase in variance may be either a cause or an effect. Mr. Maxwell himself assumes that the individual differences in actual test scores are due mainly to "genetic or hereditary causes", and consequently reflect genuine differences in innate ability. Now it is known that a common contributory factor making for maladjustment is exceptionally high or low innate ability in the child. A child who is unusually bright (and therefore quite possibly more intelligent than his parents or brothers and sisters) is apt, like Joseph of old, to become maladjusted at home. A child who is dull is apt to become maladjusted at school. Consequently we should expect a sample of neurotic children to contain a slightly higher proportion of both bright and dull children than would be found in an unselected sample.

However, Mr. Maxwell tells us that in his group the test results are "uncorrected for unreliability", i.e., the scores are taken as they stand. Now scores obtained in the WISC are bound to be affected by the emotional attitude of the child when tested; and these attitudes vary more widely and more strongly among neurotic children, particularly when the test forms part of a medical examination at a clinic, than they do among normal pupils tested merely for purposes of research. In some cases the child referred for special examination feels that his fate may depend very largely on his doing his very best; in other cases he becomes shy, frightened, or emotionally disturbed, and so fails to do himself justice. In the neurotic group therefore environmental and motivational factors may have played a larger part than usual. Additional causes may also have been at work. Dr. Jones personally tested all the children in her sample; and that may have introduced greater uniformity in administration and scoring. Mr. Maxwell's data were selected from files covering a fairly long period of years, and were therefore presumably obtained by different psychologists at different periods.

Nevertheless, even if these somewhat speculative suggestions are justified, we recognize that it still remains conceivable that Mr. Maxwell's hypothesis of assortative mating may account for part of the discrepancy he himself has noticed. For one thing his sample was obtained in a different locality from ours; and in any case the figures and calculations he has carried out are themselves of special interest. What seems really needed is a direct estimate of the correlation between fathers and mothers, based, for example, on the occupations of the fathers and the psychologist's impressions of the mothers, who usually accompany the children. Our main purpose, however, is to suggest that the reader should beware of making any hasty generalization about genetic influences before other factors have been taken fully into account.

MARGARET HOWARD
CYRIL BURT

NOTES AND CORRESPONDENCE

EXPERIMENTS ON TELEPATHY IN CHILDREN

A Reply to Sir Cyril Burt

By C. E. M. HANSEL

In his recent paper [this *Journal*, XII, pp. 55f.] Professor Burt discussed the experiments on the Welsh schoolboys presented by S. G. Soal and H. T. Bowden in their book *The Mind Readers*. The conclusions he reached were quite different to those which I myself had reached and reported in the *New Scientist*, V, pp. 983f. and the *Manchester Guardian*. Professor Burt criticizes my attitude to experiments on extra-sensory perception, my criticisms of the experiments in question, and my "armchair method" of dismissing awkward observations. Many of his statements are in my view misleading, others a matter of opinion. The final case rests on the defects of the experiments themselves; and in order to make an adequate reply and a complete defence of my criticisms it would be necessary to analyse them in some detail. However, this note¹ will be concerned chiefly with an examination of Professor Burt's criticisms and arguments; elsewhere it may be possible to give an examination of the experiments on the Welsh schoolboys.

Professor Burt starts with two statements: (i) that there is "an extremely strong case for the hypothesis of 'extra-sensory perception'", and (ii) that those psychologists who advocate a 'hypothetico-deductive' procedure as the foundation of scientific methodology "ought now to accept it as definitely proved". I do not agree with the first statement, and can find no experimental evidence to support it. I note, however, that the experiments under discussion are described by Professor Burt as follows: "for the care with which they have been conducted the experiments here recorded are unrivalled in the whole corpus of psychical research"; and again "were it not for the strong antecedent presumption *against* extra-sensory methods of obtaining factual information, his (Dr. Soal's) results would, I imagine, be hailed as virtually conclusive". In the discussion which follows I will therefore confine myself mainly to the experiments on the Jones boys, and will later show that, rather than provide a strong case, they provide a completely ridiculous case to support—let alone prove—the hypothesis of extra-sensory perception.

A psychologist who follows a hypothetico-deductive procedure needs to be as careful as a psychologist who follows any other procedure to see that an experiment is carefully designed and executed. An experimental result can only satisfactorily confirm a particular hypothesis if the experimental design excludes alternative explanations which render the hypothesis unnecessary. To follow Professor Burt's argument to the limit we should have to conclude that those psychologists who follow a 'hypothetico-deductive' procedure ought now to accept telepathy as proved by the experiments on the Jones boys, and that they should also accept the fact that the Jones boys obtained their high scores by passing

¹ Mr. Hansel has been good enough to submit with his paper a detailed appendix defending his explanations by explicit page-references to *The Mind Readers* and a full discussion of the results. This seemed tantamount to a second 'critical notice' even longer than the first; but we hope it may be published in full elsewhere. Mr. Hansel states that readers interested in the problem can obtain a copy from him; and this we certainly hope they will do (*Editorial Note*).

signals from one to the other ; for in my opinion this second hypothesis is confirmed at least as adequately as is that of telepathy.

Professor Burt classifies arguments put forward by critics as *a priori* and *a posteriori*. In the case of *a priori* arguments, I have given reasons why we have to be cautious before accepting evidence which claims to prove extra-sensory perception. Any experimental result must be treated with caution before it has been confirmed ; but a result which is at variance with existing knowledge is partly contradicted at the start, whereas a result which is in accordance with existing knowledge is partly confirmed. If an experimental result can be understood in terms of existing knowledge, we are likely to be sceptical of any new explanation which is advanced, particularly if this new explanation requires a revision of generally accepted principles. Thus I think that *a priori* arguments affect the attitude we adopt towards the experiments, and the amount of confirmation we consider necessary before we become convinced by them.

It is necessary to differentiate between the result of an experiment and the hypothesis put forward to account for it. I accept the fact that high scores have been produced in some ESP experiments and that these scores cannot be attributed to chance. I do not agree with the hypothesis of ESP which has been put forward to account for the results. Given an experimental *result* which has been adequately confirmed, a hypothesis put forward to account for that result—or put forward before the result was obtained—will not necessarily be confirmed to the exclusion of other hypotheses. It may be necessary to carry out further experiments to confirm or reject a *particular* hypothesis. In order to confirm an hypothesis involving a new process it is necessary that the experimental design should exclude other hypotheses which involve only known processes. If the Jones boys had produced high scores under conditions in which the possibility of their passing information by known methods was excluded, it would be necessary to consider the hypothesis of telepathy. We find, however, that on all those occasions when the conditions were of this nature, the boys failed to achieve high scores. Thus far from confirming the hypothesis of telepathy the results may be said to support the hypothesis that information was passed by means of signals.

The *a posteriori* arguments are those which invoke other explanations than those involving ESP to account for the experimental results. Experimenters in psychical research have clearly been aware of the fact that they must exclude alternative explanations than that of ESP. It has however been questioned whether this is possible, the argument being that a critic may say that all the experimenters are conspiring to delude the public. But the normal means of verification used in scientific investigations is for the experiments to be repeated by independent workers until this possibility becomes more and more remote. Experimenters in psychical research are well aware of the necessity for their results to be confirmed, but they have always shown the greatest reluctance to pass on their high-scoring subjects for tests to be carried out by critics. So far as I know, on no occasion has Dr. Soal sent one of his subjects to a psychological laboratory for his observations to be confirmed by other investigators. In the field of psychical research in general we find after many years of experimental work, firstly, an inability to obtain confirmation of experimental results and secondly the tendency to produce new hypotheses to explain away this inability. If we accept all these new hypotheses, verification of ESP by confirmatory tests becomes virtually impossible.

In my view *a priori* arguments determine our attitude towards an experiment, and may save time and effort in scrutinising every experiment, however fantastic it may appear. *A posteriori* arguments, on the other hand, should be anticipated when designing an experiment or be eliminated by further experimental work.

In my article in the *New Scientist* I defined the four processes : Telepathy, Clairvoyance, Precognition and Psychokinesis. I then stated that these processes conflict with what we know about :

- (1) The way information is passed from one person to another.

- (2) The dependence of mental processes on physical processes in the nervous system.
- (3) The dependence of an event on preceding events.
- (4) The laws of mechanics.

The fourth principle was intended to apply to psychokinesis but not to telepathy, just as the first principle was intended to apply to telepathy and not to psychokinesis. I thought this would be clear to a reader without further comment. These four points were not culled from any source—such as Dr. Broad's article to which Professor Burt refers in his footnote on page 60. They were merely intended to give four fairly obvious reasons for doubting the conclusions drawn from the experiments. I omitted to point out Dr. Broad's conclusion because I have not read his article. Professor Burt refers to these four points as "four 'basic requirements of science'". But I nowhere referred to them as such. He has taken his quotation from the last paragraph of my article where I say "Those who cannot afford the time to investigate the literature on the subject themselves are advised, before being swayed by such opinions to demand that just one experiment be produced which satisfies *two basic requirements of science* : first, the result must have been confirmed by independent critical investigators ; second, the experimental design must be sound".

Professor Burt asks how I can be so sure that telepathy is 'impossible'. He says that there are cases where absolute impossibilities can be affirmed in this way, on purely *a priori* grounds, and states as an example the proof that a regular septahedron is impossible in euclidean space. But, when I say that telepathy is impossible, I mean that processes do not occur in nature as we know it which permit its occurrence. In the sense that existing knowledge *may* be wrong nothing is impossible. In this sense it is *possible* that someone should construct a regular septahedron : for there may be errors, so far undetected, in geometry. In the philosophical sense *any* proposition has a certain probability of being true. But *telepathy either can occur or cannot occur*. When examining the experiments, it is not unreasonable to assume that telepathy cannot occur in view of the *a priori* arguments against it; and in fact a vast amount of the experimental work has failed to give a satisfactory demonstration of its occurrence. If, in the case of an experiment which we are examining in this way, we find our initial assumption is untenable, then we are quite free to make and other assumption to account for what happens.

I would not argue that the alternative to telepathy must always be "some kind of trick". Most of the results produced in psychical research in my opinion have been due to the unsatisfactory design and execution of the experiments. Nor can we say that the facts as recorded render this alternative (trickery) improbable. The facts as recorded in the case of the experiments on the Jones boys make it appear to me extremely likely that trickery did take place. This is because the recorded facts are easily explained in terms of trickery, and the effects on the scores following a large number of variations in the experimental conditions fit in a remarkably consistent manner with what we can predict would happen if a trick was being used. Professor Burt quotes me, on p. 99 of his paper, as saying "We know that there must be some explanation of the trick which involves only natural processes". This statement was made in my article at the end of a paragraph where I was discussing the attitude we adopt when we see a lady apparently sawn in half by a conjuror.

Many of Professor Burt's statements which he makes during his excursion into the realm of theoretical physics are due to his unfortunate misinterpretation of my statement about the 'laws of mechanics' ; and there would appear to be little point in my discussing them further. I should, however, point out that I nowhere *used* the word 'cause' in my article (see footnote to p. 61) ; but I *do*, in fact, assume that a high score in a telepathy experiment is caused by something ; and Professor Burt appears to assume that it is caused by telepathy which in turn may be dependent on "sub-atomic processes" in the percipient's nervous system.

Professor Burt speaks about causality and determinism as if they were relics of a simple cosmology "as obsolete as the bonnets and bustles of Queen Victoria herself". We should first note that both bonnets and bustles are likely to come into fashion again; but it is also doubtful if the physicists he mentions would agree with him. Planck, for example, has written "The principle of causality must be held to extend even to the highest achievements of the human soul. We must admit that the mind of each of our greatest geniuses—Aristotle, Kant or Leonardo, Goethe or Beethoven, Dante or Shakespeare—even at the moments of its highest flights of thought or in the most profound inner workings of his soul—was subject to the causal fiat and was an instrument in the hands of an almighty law which governs the world".

Professor Burt invokes processes of a "sub-atomic order" in the nervous system. But if sub-atomic processes could affect the nervous system of a person so that he thought of a particular thing (as determined by what symbol was being shown—or not being shown, in the case of clairvoyance—to some other person) why should there be this correspondence between the thought of the agent and the "guess" of the percipient if these sub-atomic processes follow no causal laws. If these sub-atomic processes enable a percipient to achieve 25 hits in 25 guesses then surely they are behaving in a remarkably consistent manner. Physicists like Jeans, Eddington and Schrodinger who have speculated on whether mental processes can be affected by the unpredictable behaviour of a few atoms in the brain, have had to discard the idea.¹ Eddington says, "If I have at all rightly understood the processes of my own mind, there is no finicking with individual atoms": and again "It is one thing to allow the mind to direct an atom between two courses neither of which would be improbable for an inorganic atom; it is another thing to allow it to direct a crowd of atoms into a configuration which the secondary laws of physics would set aside as 'too improbable'".²

In his postscript, Professor Burt writes: "At that time Mr. Hansel believed that 'possibly never again would a percipient be produced who can consistently obtain high scores'". I still believe that. The Jones boys had, according to information available to me, stopped producing high scores some time before I wrote my article, and they have not yet been produced (since they recovered their powers) for testing by critics.

Professor Burt says, in relation to the Jones boys: "Accordingly, as in this case his (Hansel's) earlier explanation was ruled out by the procedure adopted, he has since put forward . . . two new suggestions to account for their unexpected achievements". The 'earlier explanation' to which Professor Burt refers is presumably my explanation of the Shackleton experiments. This explanation was not intended to apply to every other experiment carried out in psychical research. I have not produced *alternative* explanations in regard to the experiments on the Jones boys: but I have tried to show that these boys used more than one trick. The answers to Professor Burt's other points are covered in my account of the experiments; but it is necessary to answer his accusation that I have dismissed awkward observations.

In my view, experiments in psychical research have established just two things quite clearly and definitely: (1) subjects sometimes obtain above chance scores which cannot be attributed to chance, and (2) subjects sometimes indulge in trickery. Both these points are admirably demonstrated in the experiments with the Jones boys, as no one is likely to dispute the significance of their scores, and they admitted on two occasions to having cheated.

When he refers to 'awkward observations' does Professor Burt mean the scores—which I accept—or the hypothesis of telepathy which has been put forward to account for these scores? I reject this hypothesis just as Professor Burt rejects my 'imaginary' explanations.

¹ J. Jeans, *The New Background of Science* (1933), p. 280.

² A. Eddington, *The Nature of the Physical World* (1928).

EXPERIMENTS ON TELEPATHY IN CHILDREN

A Reply to Mr. Hansel's Criticisms

By CYRIL BURT

Common Ground. There is, I fancy, far less difference between Mr. Hansel and myself than he appears to believe, since many of the views which he so vigorously attacks are views which I should myself reject as emphatically as he does. Nevertheless, on two or three points, it is clear, we disagree profoundly. Moreover, these relate, not merely to alleged telepathic phenomena, but to the far wider problems of psychological method and hypothesis. Partly for this reason, and partly because his conclusions have been so widely quoted, I shall try to answer his criticisms as explicitly as I can.

As it stands, Mr. Hansel's opening paragraph gives the reader a very misleading impression of my views. "Professor Burt", he says, "starts with two statements: (i) that there is an extremely strong case for the hypothesis of 'extra-sensory perception', and (ii) that those psychologists who advocate a hypothetico-deductive procedure 'ought now to accept it as definitely proved'." Mr. Hansel then tells us that he "does not agree with the first statement". Neither do I. In point of fact I started with no such statement. After summarizing the 'Methods and Results' of Dr. Soal's latest experiments, I went on to say that "*prima facie* these results *seem* to establish an extremely strong case for 'extra sensory' perception". By omitting the first seven words Mr. Hansel converts my preliminary indication of the conclusion ostensibly suggested by Dr. Soal's work into a generalized postulate of my own. But I immediately added that "nevertheless, the majority of scientists and present day psychologists" (among whom I included myself) "will hesitate to accept Dr. Soal's conclusions without closer scrutiny of the evidence"; and I pointed out that "there are several manifest loopholes on which a sceptic could seize". After a rather lengthy argument I myself concluded that, in spite of the striking character of the results obtained, "an impartial critic is bound to insist that a single research by a single pair of investigators on a single pair of boys must be further confirmed by other investigators, and a working hypothesis must be elaborated to coordinate the new data with accepted theories (modified if necessary) and itself be capable of verification: *only then*", I argued, "can the results be accepted at their face value". So far therefore we are on common ground.

In re-phrasing the second of the two quotations Mr. Hansel again saddles me with a doctrine that I expressly criticized and indeed disowned. "To follow Professor Burt's argument to the limit", he says, we could just as well claim that "the Jones' boys obtained their high scores by passing signals from one to the other, for this hypothesis is confirmed at least as adequately as that of telepathy . . . An experimental result can only satisfactorily confirm a particular hypothesis if the experimental design excludes *alternative* explanations which would render the hypothesis unnecessary"; and that requires that the experiment should be carefully planned with this aim in view. But this is not "a reply"; it is almost a verbatim echo of my own contentions.¹ If Mr. Hansel will read on a little further he will

¹ In case the reader does not have the passage in front of him may I quote it here? "The hypothetico-deductive method as ordinarily stated", I wrote, "is fallacious in itself and inconclusive in actual practice: it implies that, in solving any particular problem, only one hypothesis calls for testing, whereas two at the least are essential—namely, the hypothesis to be established and its opposite or else the next most plausible alternative. In either case rebuttal of the alternative is vital. And this", I added, "implies that the plan of experiments must be carefully designed before the observations are actually made". At the same time, I insisted, much as Mr. Hansel has done, that if some relatively new hypothesis is to win acceptance, it "must not stand in chilly isolation, but be fitted into the context of current scientific opinion" (this *Journal*, XI, 1958, pp. 59f).

find a section headed "Alternative Hypotheses". Those numbered i, ii, *a*, *b*, *c*, and iii between them cover all the methods of "passing signals" which Mr. Hansel mentions; and I briefly indicated what were the experimental conditions which seemed to exclude them as possible explanations (*loc. cit.*, pp. 66-68; cf. also pp. 98-99).

A Posteriori Arguments. Mr. Hansel next deals with my discussion of *a priori* and *a posteriori* arguments. To the latter phrase he gives an interpretation which differs from that of most other methodologists. "*A posteriori* arguments", he says, "are those which invoke other explanations than those involving ESP to account for the experimental results". The explanations on which he himself lays most stress may, I think, be summarized under three main heads.¹ As he apparently thinks that in my previous paper I rejected these particular explanations a little too readily without adequate examination, it will be necessary to consider each of them in greater detail here. At the same time I shall try to answer his final query: what did I mean by the 'awkward observations'² which, I complained, he himself had 'dismissed'—or (as perhaps I should have said) 'ignored'.

(i) "We do not", he says, "commence with the assumption that the same trick was employed over the full period of the experiments". He thinks that at the start, when the Jones boys sat at the same table, they used "tactual cues"—the agent nudging or tapping the percipient with his foot or leg. When, as in the most of the experiments forming the main series, the distance between them was too great for signalling in this way, he believes they changed to *visual* cues, still given by movements of the foot or leg. Owing to inadequate screening, he supposes, the percipient could have readily seen signals of this kind; and, if the movements were very slight or very slow, then, he maintains, they would have escaped the notice of the independent observers who were instructed to look out for such signals. To this there is a simple reply. A more careful study of Dr. Soal's book shows that high scores were made (*a*) on several dark evenings when such movements could not have been seen, (*b*) when Glyn (the percipient) was so seated that he could not have watched Ieuan's legs without a very obvious movement of the head, and later on (*c*) when it became possible to provide quite efficient screening. Accordingly on these occasions Mr. Hansel assumes that the visual cues were signalled by an older member of the family (either the father or the older brother Gareth) who was stationed at one of the upper windows in the house, "using binoculars if necessary". This is surely rather far fetched. But in any case, as Dr. Soal himself has pointed out, some of the best runs with high scores during daylight hours were obtained when both the father and the brother were away at work.

(ii) Consequently, Mr. Hansel now seems rather to favour *auditory* cues. "The most likely type of signal in view of the experimental conditions" (he says) "is some sort of high pitched whistle"—a supersonic instrument, inaudible to most adults, "blown by pressing a rubber bulb concealed by sitting with folded arms". Such whistles, he says, are "well known in the Snowdon area where they are used for controlling sheep dogs".³ His

¹ Mr. Hansel's explanations were briefly described in a lecture to the Society for Psychical Research in the autumn of last year: and have been amplified further in one or two later papers: see *J. Soc. Psychic. Res.*, XL, pp. 291-9 and refs.

² Mr. Hansel asks if by this phrase I meant "the scores" or "the hypothesis of telepathy". I meant neither. I meant the detailed observations reported on those occasions in which special precautions were taken to exclude trickery of the kind he has in mind—e.g., in the successful experiments carried out on August 10, when the boys "wore close fitting bathing trunks and were carefully searched" (mentioned in a paragraph of my article just before the one he quotes) and other 'observations' of the same sort described below.

³ This remark seems rather an exaggeration. During several years' stay in North Wales, as well as on occasional visits to the Snowdon district, I have never encountered a shepherd who used such a thing; and very few knew of the existence of *ultrasonic* whistles.

suggestions have been supported by Mr. Christopher Scott and Mrs. Goldney in a joint article published quite recently.¹ Both they and Mr. Hansel have carried out experiments to illustrate their arguments. (a) Most of Mr. Hansel's experiments were carried out with an Edelmann whistle. This is a bulky and fairly expensive instrument, and, so far as is known, not on sale in this country. It is therefore difficult to conceive how such a thing could have been procured by the boys. (b) Mr. Scott and Mrs. Goldney carried out their experiments with a Galton whistle which is far less easy to use.² It can, however, be purchased from C. F. Palmer (Scientific Instrument Makers, Effra Road, Brixton). Inquiries indicate that no purchase has been made through them which could be attributed to the Jones, even if we imagine anyone in a remote Welsh village had ever heard of such a thing or had been willing to invest four guineas in the hope of baffling Dr. Soal and his collaborators. (c) Dr. Soal, I believe, was the first to point out that a high-pitched dog-whistle had at one time been on sale at a shop some seven miles away from where the Jones lived³; and Mr. Hansel now seems to believe that this was the instrument used by the boys. The whistle is blown by the mouth, and its highest pitch is still audible to the vast majority of adults under 50. We should therefore have to suppose that the Jones family, after hearing of Galton's discovery, proceeded to adapt the whistle by the methods described by Mr. Hansel so as to obtain a higher pitch, and fitted it with a bulb which would produce a strong but inaudible puff. This is by no means an easy operation even for a technician in a psychological laboratory working under the instructions of a psychologist who is familiar with all the details of auditory thresholds at various ages.

However, Dr. Soal's recent experiments seem almost conclusive. He found that Glyn could just hear the Galton whistle when set to a length of 3 mm, provided the distance was not much above 100 feet (i.e., much less than the distance at which they obtained good scores). On the other hand, he found many adults of about 30 could hear it quite as well, including Mr. Owen who had attended several of the original experiments, and stood close by the agent who is supposed to have used it.

A final refutation of the theory is to be found in the successes achieved when the boys, besides being searched, were required to wear bathing costumes. Neither Mr. Hansel nor Mr. Scott refer to the results obtained on these occasions.⁴

(iii) Presumably to meet this type of objection, Mr. Hansel brings forward his last alternative, namely, that, when the boys were too far apart for the percipient to detect any supersonic signals transmitted by the agent, then one of the older Jones's might have used it to facilitate the cheating. Under these more difficult conditions, we are told, "high

¹ 'The Jones Boys and the Ultrasonic Whistle', *J. Psychic. Res.*, XL, pp. 249-59. The reader should note that on a later page of that journal Mrs. Goldney, who, as she says, "witnessed very high scoring on so many occasions", states that she herself endorses the view expressed by Professor Mundle: "I do *not* believe the boys were using a whistle"—an opinion (she adds) "which, I believe, is shared by most, if not in fact all of those who repeatedly saw the high scoring" (p. 272). Surely this statement from an impartial, highly trained, and highly critical first-hand observer must strongly outweigh the rather biased guesses of others who did not witness the experiments.

² For a detailed discussion of these earlier instruments see Titchener, *Experimental Psychology*, Vol. II, pt. ii, pp. 31f. One difficulty is that the pitch varies not only with the adjustment of the tube but also with the pressure applied. Further, unless great care is taken, the puff of wind produced by the bulb is readily audible even when the whistle is not.

³ Actually this shop had been closed for some time when Dr. Soal's experiments were carried out; and his own inquiries indicated that such whistles were not in fact used by shepherds in that part of Wales. In any case, the Jones family belonged to the mining community, not to the sheep-farmers, as most of the critics seem to have assumed.

⁴ In an appendix on "Conditions under which high scores were not obtained" Mr. Hansel refers (by date) to an incident on 9 August when, as he says, "the boys were asked to perform in bathing trunks, and Glyn said he was not going to try" and accordingly produced nothing but random guesses. But this reference as it stands is very misleading. Glyn's failure to respond was due, not to the fact that both were asked to perform in bathing costumes, but (as the book explains)

scores were only obtained when a member of the family was present". In point of fact this is by no means correct: high scores were frequently obtained when no other members of the family were present, e.g., on August 10 (as mentioned above) and on several other occasions. Even when one of the senior members was invited to assist, this was often arranged at a moment's notice, so that little opportunity was given to prepare instruments and other details.

In the foregoing paper (p. 176) Mr. Hansel complains that "experimenters in psychical research have always shown the greatest reluctance to pass on their high-scoring subjects for tests by critics". "So far as I know", he says, "on no occasion has Dr. Soal sent one of his subjects to a psychological laboratory for his observations to be confirmed". It is, of course, by no means easy to "send" other people to distant laboratories. But as a matter of fact, Dr. Soal did arrange for the boys to be tested both in the Psychological Laboratory at Birkbeck College and at the rooms of the Society for Psychical Research. During some of these sittings Dr. Soal himself was absent. Professor Mace has stated that he saw no signs of cheating, and so have all the other observers. Moreover, on 2 August, 1958, after arrangements had been made over the telephone by Dr. Soal, Mr. Hansel himself, with his niece and Mr. and Mrs. Scott visited the Jones family in Wales. At the same time they made inquiries in the locality about inaudible sheep-whistles; but the whistles obtained, says Mr. Scott, were "far from ultra-sonic". Before any experiments were carried out the boys' pockets, etc. were examined. However, the final results, so Mr. Scott assures us, "were devoid of interest". (It might perhaps have been wiser to follow Dr. Soal's procedure and carry out the search *after* the scores had been made.)

After this it was decided to arrange "a situation which would offer the highest motivation" and at the same time provide an opportunity for introducing (unknown to the boys or their family) an electronic ultra-sonic sound-detector. It was agreed with Dr. Soal that there was "one ideal solution—a television programme". During the previous October the boys had already given a successful demonstration on television from the Manchester studios: the odds against the scores then achieved were 200 to 1. And, from the keenness they displayed, it seemed clear that an exhibition of this kind would furnish an even greater inducement for the boys to do their best than the mere monetary rewards. A new performance was accordingly arranged with the London producers of 'Panorama'. The boys appeared highly confident of success. Nevertheless, no high scores were achieved, and certainly no attempt to use ultrasonic signalling was detected by the oscillator. After all this, it hardly seems fair to suggest that Dr. Soal "showed the greatest reluctance" to pass on his high-scoring subjects. The plain fact is that Mr. Hansel has apparently done his best to obtain evidence for the use of the devices he postulates, and has so far failed. These attempts are nowhere mentioned in his present paper or in the appendix.

One experiment seems of itself to provide the strongest evidence against Mr. Hansel's hypothesis of a prearranged clues. "On 7 April 1956", says Dr. Soal, "five new coloured pictures were sprung upon the boys without a moment's warning". In the first set of runs Glyn scored 47 successes out of 100 (the expectation being 20); the distance was then quadrupled, and there was no significant drop (43 out of 100). The odds against this

to the fact that Mr. Bowden had brought a bathing dress for Ieuan and left Glyn "to sit in his own threadbare and shabby costume". Mr. Hansel remains silent about the high scores obtained on August 10 at distances of 160 feet, when Mr. Bowden had also brought a suit for Glyn, and Glyn was once more cooperative.

My own suggestion was that, *if* any auditory clues were used, they would more likely have been supersonic signals produced by "tongue-whistling", with no apparatus whatever (*loc. cit.* p. 67). However, a good deal of highly intelligent practice with the adult observer is necessary before one can hit on the precise pitch which will be audible to a given boy of 12 or 13, but inaudible to some particular adult; and I should think almost impossible for a country lad to determine, and keep to, a note which would be inaudible to a *number* of young adults. Certainly neither Mr. Hansel nor Mr. Scott have produced any evidence to show that this would have been really practicable,

being due to chance are " 10^{15} to 1 at the least". As Dr. Soal observes, "the *immediate* success of this experiment makes it difficult to believe that a code could have been employed".

For Mr. Hansel, who now has in front of him Dr. Soal's account of the conditions imposed on this occasion or that, it is comparatively easy to think of devices by which those conditions might have been defeated. But the Jones family had no such detailed information supplied in advance. And yet, according to Mr. Hansel, they must have been prepared to adapt their varied forms of trickery to each new set of restrictions as they were "sprung upon them", and also to guess what particular loopholes would be left for them to exploit. Thus, on certain occasions they were ready with a tactile code, on others with a visual code, and on others again with an auditory code and the necessary apparatus. When in the earlier experiments out of doors, the distance was increased so as virtually to exclude signalling between the boys, the father was secretly stationed at an upstairs window, all ready with his binoculars to "relay information from the agent to the percipient"; and, when during the open-air experiments in London this device was out of question, the elder brother was ready with a practicable code, guessing beforehand that Mr. Fisk might suddenly ask him to deputize. And yet another code was available in April, when Dr. Soal came secretly provided with an entirely new set of cards. In the Manchester studios, where no electronic detector was used, they presumably signalled with their whistle; in the London studios, they somehow guessed that a detector would be secretly used for the first time in the whole series, and so wisely refrained. Surely all these highly successful anticipations would themselves presuppose on the part of the Jones family a degree of precognition or telepathic insight which exceeds anything Dr. Soal and Mr. Bowden have claimed.

These then are Mr. Hansel's main '*a posteriori* arguments'. In spite of his inquiries on the spot, no evidence is adduced to suggest these ingenious conspiracies were in keeping with what is known of character, intelligence, and scientific information displayed by the boys and their family; nor has any evidence been discovered to show that they ever possessed the necessary apparatus.¹ Noone, I imagine, would regard such a farfetched and ill supported hypothesis as in the least plausible, unless he were already firmly convinced (a) that the only alternatives² are either telepathy on the one hand or trickery on the other, and (b) that telepathy is impossible.

A Priori Arguments. This brings me to what Mr. Hansel and I call '*a priori* arguments'. It is these that have led him from the very outset to the firm conclusion that "telepathy and precognition are impossible". In my previous article I inquired (as he observes) how he could be so sure, without any first hand investigation of the facts alleged, that such occurrences were, not merely *improbable* (with which I should agree), but altogether *impossible*. His answer is that "the processes which occur in nature as we know it do not permit its occurrence".

But what precisely are these processes? And in what way do they exclude or forbid all possibility of telepathic communication? Here Mr. Hansel becomes a little vague. He repeats his enumeration of four types of process, and refers to "what we know about

¹ After the London programme was over, Mr. Scott showed the boys and their fathers the whistle he himself had used; and all declared they had seen nothing like it. The reader should remember that the experiments with the Jones boys do not by any means stand alone. I say nothing of the investigations by Rhine and his co-workers of which I have only indirect knowledge. But the striking results of Gilbert Murray's experiments with members of the Haldane family and others at Oxford have been fully described both by Murray and his cooperators. Does Mr. Hansel suppose that Murray too was guilty of trickery?

² Mr. Hansel mentions a third possibility, namely, that the design and execution of such experiments is in many cases so unsatisfactory that chance is not really excluded. But in the present case he accepts the fact that "the high scores cannot be attributed to chance."

them". I asked that, instead of headings, his 'basic principles' (as I called them) should be restated in the form of explicit propositions, so that we could know whether they were intended as postulates or as empirical generalizations. After all such phrases as "the dependence of an event on preceding events" express, not so much "obvious" principles which are "incompatible" with telepathy, but rather problems to which different scientists give different answers. What Mr. Hansel's answer is we are still left to guess from indirect allusions.

However, Mr. Hammond, who has recently come to his support (this *Journal*, XIII, pp. 92f.), makes the position a little clearer. There is in his view a single comprehensive science of the universe, with which all psychological hypotheses must conform, namely (as he puts it) "Newtonian science, in Mr. Taylor's phrase, or the laws of mechanics, in Mr. Hansel's phrase". "I take it", he adds, "that Mr. Taylor and Mr. Hansel would agree".¹ And, since Mr. Hansel has not repudiated Mr. Hammond's interpretation of his views, we may, I imagine, assume that it expresses the real basis of his *a priori* arguments. The position is quite familiar; it is one that has been adopted by most behaviourists and physicalists, who would, of course, fully agree with Mr. Hansel that telepathy must consequently be impossible.

The reason why so many psychologists adopt these mechanistic postulates seems clear enough. They are anxious to avoid what they picturesquely call a "ghost in the machine", i.e., to avoid anything like a dualistic or a pluralistic universe. But Mr. Hansel's emphasis is on what we *know* of nature. And in reply it may at once surely be said that, so far as current knowledge takes us, the universe is irreducibly pluralistic. All we know of 'nature' consists solely of interactions; and our knowledge is itself the end result of a complex chain of interactions. Moreover "the processes which occur in nature as we know it" consist, not of one type of interaction—namely, that which conforms to the 'laws of mechanics'—but of at least four types; and, except for the empirical fact that certain entities happen to take part in more than one type, these interactions are themselves quite unrelated to one another, and, so far as we can see, incapable of being reduced to a single type. Among physicists this is by now an acknowledged truism. But, since it appears to be unfamiliar, or at least unacceptable, to those who favour Mr. Hansel's general *Weltanschauung*, it is perhaps desirable to explain it at greater length.

(i) The mechanical universe, as conceived by Newton, or rather in the Newtonian cosmology as outlined by Mr. Hammond, is a universe of medium-sized objects, such as can be seen by the naked eye, interacting with one another in virtue of the force of gravitation. This is in point of fact the weakest of the interactions studied by modern physics. However, its effective strength varies with the mass of the body which exerts it. It thus enables us to account for the most familiar of the phenomena of every day observation—the falling of the apple, the revolutions of the moon and of the earth and other planets—simply because the size and mass of the two most conspicuous objects in that universe—the earth and the sun—are relatively enormous.

(ii) During the last half century or so, however, we have become increasingly aware of an entirely different type of interaction—that associated with electric charges. It is true that both the electrical interactions and the interactions due to gravity have one law in common: both obey the principle of the inverse square. But they differ in one fundamental feature. Electrical forces may produce either attraction or repulsion; and in certain circumstances the two opposite tendencies may neutralize each other. Gravitational forces (so far as we know) are responsible solely for attraction, and therefore cannot be neutralized.

¹ Mr. Hansel's later comments do in some measure seem to qualify what originally appeared to be an unreserved denial. He now explains that he meant his statements about the 'laws of mechanics' to apply to psycho-kinesis rather than telepathy. But this reservation is itself a little puzzling. Telepathy, he says, conflicts with what we already know of the way information can be passed from one person to another. But unless that means that we know the process to be wholly mechanical, there would be no conflict.

Most of our household objects—the tables, the chairs, the books, and the crockery—are electrically neutral ; but they all react to the force of gravity. The dining-room table rests heavily on the floor ; the jug rests on the table ; if the jug slips off, it falls to the earth in virtue of the earth's gravity. But the force of gravity is too weak for the attraction which one household object exerts on another to be discernible by ordinary observation. Both the cricket ball and the wicket possess mass, and must therefore attract each other ; yet no bowler would rely on the gravitational attraction of the wicket to direct his ball, although he skilfully allows for the gravitational attraction of the earth in determining its trajectory.

Now, since there is no known connection between gravity and electricity, it would be quite easy for a fanciful physicist to conceive of two completely separate universes existing simultaneously—(a) one consisting of elementary particles, which, like the cricket ball, have mass (and therefore gravitational interaction) but no electrical charge (and therefore no electrical interaction), and (b) a second universe consisting of elementary particles, which, like the photon, are subject to electrical interaction, but have no mass, and are therefore not subject to gravitational interaction. In that case the two universes might actually interpenetrate each other ; and yet an observer belonging to the electrical universe would be quite unaware of the simultaneous existence of the other universe, while a second observer, whose body was subject solely to gravity, would remain wholly ignorant of all electrical interactions—as was indeed virtually the case with human observers until comparatively recent times.

However, in the actual universe as we know it, certain entities—electrons and protons, for instance—are subject to both kinds of interaction. Hence the two sets of interactions (for reasons which we cannot even guess) are in some measure linked. And the earliest observers insisted, as Mr. Hansel does, that there could be only one type of interaction—that which obeyed the laws of mechanics—and consequently supposed that electrical manifestations were due to the mechanical effects of hypothetical 'fluids'.

(iii) Certain entities, however, for example the neutrino, appear to have neither mass nor electric charge. As a result a neutrino may pass right through the earth or sun without interacting with either. At this moment about a hundred million of the innumerable neutrinos which pour out of the sun are passing through your head every second. We might therefore feel tempted to suppose that a universe, constructed solely of neutrinos, might actually exist about which we could know nothing. In point of fact, however, it is now known that neutrinos may interact with other particles in virtue of what is called 'weak' interaction. Consequently one neutrino in about a thousand million will exhibit a slow interaction with the earth as it passes through.

(iv) The distinctive properties of the atom, however, depend upon its nucleus. And the particles within the nucleus are held together by a yet fourth type of interaction, namely, the so-called 'strong' interactions, which until quite recently no scientist imagined could ever be broken up. To do so the modern alchemist has to use energies a million times greater than those required to break up a molecule into its constituent atoms.

In the sub-microscopic world of elementary particles the effects of the gravitational forces, which Newton invoked to unify the larger cosmos of his day, are so feeble as to be almost negligible. Even the 'weak interaction' of the neutrino is stronger than the interaction due to gravity by a factor of about 10^{27} . The interactions between neighbouring atoms, and between the electrons and the nucleus within the atom, which are due primarily to electrical forces, are ten thousand million times stronger than the 'weak' interactions, and therefore incomparably stronger than gravity. Moreover, they act, not in accordance with the laws of Newtonian mechanics, but in accordance with the very different laws of quantum mechanics. If the forces acting on the electron obeyed Newtonian laws, it would spiral into the nucleus, and the atom as such would survive only for a fraction of a second. It is chiefly in virtue of these electrical interactions that our eyes react to visual stimuli, and our nervous system reacts to the impulses reaching it from our various sense organs. How such interactions in turn set up that unique psychical reaction which we call awareness

still remains an unsolved enigma. The process plainly falls outside the Newtonian system described by Mr. Hammond.

So far then from one set of principles or laws governing the entire aggregate of observable phenomena, as Mr. Hansel and Mr. Hammond suppose, we have a number of independent interactions, only partly overlapping ; and the principles which they obey, as empirical investigation shows, may vary quite unexpectedly according to the class of phenomena to which they relate. But what seems still more important for the psychological theorist to bear in mind is this: of the interactions mentioned, some were *entirely unsuspected* until comparatively recent times. The textbooks, which Mr. Hansel and I studied in our school days, contained no mention of the 'strong' interactions which, as we have seen, are the 'strongest' of all ; and the discovery of 'weak' interactions is still more recent.

That being so it remains conceivable that yet another universe and yet another type of interaction might exist, awaiting more adequate investigation, namely, a 'psychic' universe, consisting of events or entities linked by 'psychic' interactions, obeying laws of its own, and interpenetrating the physical universe and partly overlapping it, much as the interactions already discovered and recognized overlap with each other. Such a set of interactions might therefore account for the apparently super-physical phenomena of telepathy, clairvoyance, extra-sensory perception, and the like, *if* of course these phenomena could be conclusively established.¹

Such an assumption is not so far fetched as it might seem: for we have direct experience of what we loosely call our sensation, thoughts, and feelings—the colours, sounds, smells, ideas, beliefs, pleasures and pains, and all the forms of consciousness, which a purely behaviouristic psychology tends to ignore ; these are certainly not included among the events and entities studied by the physical sciences ; and yet they appear to be partly dependent² (in ways we do not understand) on the physical processes taking place in our brains or nervous systems.

Telepathy and Subatomic Processes. So far all I have attempted to show is that the *a priori* arguments in no way compel us to deny the possibility of telepathy in advance. I have not, either here or in my previous papers, sought to maintain that telepathy is "proved", much less that we have any satisfactory explanation of the phenomena alleged. Mr. Hansel, on p. 177 of his 'Reply', declares that "Professor Burt appears to assume that

¹ To avoid misunderstanding, let me say in passing that I do not regard these generalizations as necessarily final in their present form. I look on the physicist's accounts of the different interactions I have mentioned as analogical models rather than literal descriptions of what is actually taking place ; and as models they may be merely provisional—the most satisfactory interpretations that have been worked out up to the present, not necessarily the best that might conceivably be devised. Consequently, I would guess that, when increasing knowledge yields some more plausible explanation of the facts of sensory and extra-sensory perception, it may, in so doing, entail a complete revision of our present conceptual framework—a revision as radical as that already introduced during our own life time by the theory of relativity and the quantum theory : my conjecture is (as I indicated in earlier papers) that the change will take the form of some further development of a field theory, embracing both physical and psychical phenomena. This is one of the reasons why I have so insistently pleaded for further research rather than for a summary dismissal of such phenomena as impossible.

² Mr. Hansel himself refers to 'mental processes' and their 'dependence' on the physical processes in the nervous system. By 'dependence' I merely mean *functional* dependence. Mr. Hansel, however, apparently means *causal* dependence of a deterministic kind. Here the chief difference between us emerges in Mr. Hansel's footnote on the 'uncertainty principle'. Contrary to most physicists, Mr. Hansel rejects the principle and apparently insists on relations that are completely deterministic. Were we to accept the uncertainty principle, he says, it would become impossible to predict a future event. Admittedly any *complete and precise* prediction would be out of the question ; but we could still make partial, probable, or approximate predictions, and that is the most that is possible in those areas of physics in which quantum theory is essential.

it" [a high score in a telepathic experiment] "is caused by telepathy, which in turn may be dependent on subatomic processes in the percipient's nervous system". This, however, rests on a complete misunderstanding. Indeed Mr. Hansel himself, somewhat inconsistently, proceeds almost immediately to criticize me for *rejecting* all attempts to apply the concept of causation.

In the following paragraph he goes on to say that I "invoke physical processes of a subatomic order" to account for telepathy; and inquires how subatomic processes taking place within the brain could "affect the nervous system of a person so that he thought of particular things . . . if those processes followed no causal laws". This is rather like asking how the subatomic processes taking place within a milligram of radium could "affect" certain atoms so that within 1,600 years half of the radium will have disintegrated "if these processes followed no causal laws". The answer of course is that the concepts Mr. Hansel uses—one process 'affecting' another and both processes obeying 'causal' laws—are concepts which no physicist would ever think of employing when describing such phenomena.¹

But in point of fact I nowhere 'invoked' subatomic processes to account for telepathy. What I said was that, "since the physical processes in the nervous system to which he [Mr. Hansel] refers depend in part on processes of a subatomic order, we can no longer insist that the 'mental processes' associated with them conform with 'laws of mechanics', when these laws fail to hold good within the atom". In other words, my argument was, not that the subatomic processes "*affect* the nervous system" so as to produce telepathic experiences, but merely that they do not *prevent or conflict* with the conceivable occurrence of such experiences. I might equally well ask Mr. Hansel how on his assumption he supposes "physical processes of a subatomic order", or, for that matter, physical processes of whatever order, could of themselves "affect the nervous system of a person" *so that he thinks or perceives anything at all*? Does he really believe that the assumption of causality makes it any easier to account for the occurrence of even the most familiar conscious processes?

The Need for Further Research. Mr. Hansel tells us that the results of his *a priori* arguments "may save time and effort". And I suspect that this is one of the strongest motives—although perhaps an unconscious motive—in the minds of those who have so eagerly endorsed his views. Once we have persuaded ourselves that "telepathy and precognition are incompatible with science" and therefore "impossible" (*New Scientist*, V, p. 713), we need no longer bother to plan and carry out elaborate researches to test its possibility, or even to "scrutinize the experiments" of those investigators who claim to have obtained positive results.

Mr. Hansel confronts his critics with a simple dilemma. "Telepathy", he says, "either can occur or it cannot". But, since there is no general agreement as to what telepathy implies or how the processes alleged can operate, it is impossible to give a simple answer. Exactly the same type of argument was used against the early members of the Society for Psychical Research, when they proposed to investigate the phenomena of mesmerism. "Either mesmerism can occur or it cannot"; nor was it difficult to persuade the majority of scientists and doctors of that day that mesmerism, as the term was understood both by its popular supporters and by its scientific opponents, could not possibly occur; hence, it was contended, it would be a waste of time, effort, and money even to start any first hand inquiry into the subject. Nevertheless, though 'mesmerism' as commonly

¹ Mr. Hansel quotes Jeans, Eddington, Planck, and Schrödinger to refute what he supposes to be my version of subatomic processes. The first three are dead; and the passages cited from their writings were published during the very early stages of quantum theory, twenty or thirty years ago. The reference to Schrödinger I have failed to trace: it hardly agrees with what Schrödinger says about "the spiritual bearing of science on life" (*Science and Humanism*, pp. 1f.). But in any case none of Schrödinger's remarks seems relevant to our present problem.

interpreted did not occur, certain exceptional and rather neglected phenomena, which had occasioned the belief, undoubtedly did occur, and proved well worth the trouble devoted to the researches. One of the most significant results was a new interpretation of hypnotic phenomena, which today are no longer regarded as mysterious or paranormal.

So, I imagine, it will be with 'telepathy'. I should be inclined to agree with Mr. Hansel that 'telepathy', as that word is ordinarily interpreted and as he himself defines it, does not occur. Indeed, I said so in the article he criticizes (p. 81). His underlying mistake, as it seems to me, is to treat what are called 'extra-sensory' forms of cognition as something essentially incongruous and anomalous, standing apart from all other modes of cognition. But in any case I shall not seek to emulate his bold example, and pronounce in advance what can occur and what cannot. All I think we can safely say is that Dr. Soal's experiments have rendered it far more *probable* than before that something rather like so-called telepathy may occur. But, as Dr. Soal himself has emphasized, the results so far obtained have not done much to elucidate the intrinsic nature of the processes at work; nor indeed was that his primary object. This should now, in my view, form the object for the next stage in such researches; for, until we have achieved a clear conception of what it is we are trying to establish, it will be impossible to arrive, even tentatively, at a satisfactory decision. The real reason why so many sceptics prefer the hypothesis of trickery—however ill supported—is that we all of us understand, or think we understand, how trickery might work, but no one as yet understands how 'telepathy' could work. The proper way to approach such a question is not to deprecate the time and effort spent on the problem, but to undertake further and better planned investigations in the hope of securing some such understanding.

SCIENTIFIC METHOD : A REPLY

By J. G. TAYLOR

There are three major misunderstandings in the criticisms that have been levelled against the views put forward in my paper (*Brit. J. Psych.*, XLIX, pp. 106 f.). First, although I said I was concerned solely with the problem of explaining *conscious* phenomena, my critics assume that I was talking about the *whole field* of psychology. I expressly stated that I was excluding behaviour from my definitions of 'mind' and 'psychology'.

Mr. Pilkington notes the exclusion of behaviour; but is puzzled by my claim that there can be no scientific explanation of conscious phenomena apart from the science of behaviour. He thinks I have contradicted myself. But there is no inconsistency. I merely said that, if the *dependent* variables are conscious phenomena, and the *independent* variables are behavioural, then any equations we may write to express a functional relation between the two must be consistent with behavioural science. This implies no retreat from my decision to confine *dependent* variables to conscious phenomena.

One instance of this misunderstanding is the assumption that, when I referred to the nativist-empiricist controversy, I was talking about nature and nurture. I admit that I should have been more explicit. I was really referring to the question whether conscious phenomena are determined by the operations of a Cartesian 'soul' or by interaction between the organism and the environment. Had I intended to refer to the nature-nurture controversy, I should have said so. But in that case I should not have put forward the absurd beliefs attributed to me. Their very absurdity sufficiently shows that my critics have misinterpreted my argument. Mr. Pilkington supposes that I should deny that pitch discrimination requires both a normal cochlea and training. If he had been more consistent in his misreading, he would have concluded that I should maintain the still more absurd belief that pitch discrimination requires no cochlea at all: for, if he substitutes 'nature' for 'nativist', and 'nurture' for 'empiricist', then my acceptance of an uncompromising empiricism would for him imply that I hold that the mind has nothing whatever to do with the body. Logically his misunderstanding leads to the conclusion that I accept what Professor Ryle has satirized as 'the dogma of a ghost in the machine', which of course was far from my intention.

The second misunderstanding is related to the first. Because they fail to note my limitation of the universe of discourse, and assume that I was discussing psychology in general, my critics have missed the meaning of my argument that statistical studies are not likely to lead us to a satisfactory theory of conscious phenomena. Since consciousness occurs in an individual, and there is no analogous sense in which we can speak of a 'group mind', the laws of mind must be expressed in such a manner that the conscious experience of a given *individual* can be deduced, provided of course we can assign values to the independent variables. My analogy of the speedometer needle seems to have been overlooked.

The extent of the misunderstanding is brought out in the argument used by Professor Burt and Mr. W. Gregory (this *Journal*, XI, p. 122). Having argued that a Newtonian type of theory is inapplicable in psychology because it requires measurements to be made in terms of absolute units, they illustrate their point by considering an equation of Galileo's, relating to the movement of a ball rolling down an inclined plane, viz., $s = \frac{1}{2}at^2$, where s is the distance travelled by the ball at time t after its release, and a is acceleration due to gravity. "If", they say, "the variables were psychological measurements, such a treatment would rarely be possible. Suppose, for instance, that s represented the positions

of pupils assessed for their progress in arithmetic, and t their rank-order for (say) industry, and that we then wanted to express the relation between the two in a Galileian formula, there would be no means of discriminating between $s = \frac{1}{2}at$ and $s = \frac{1}{2}at^2$; the most we could do would be to say that the two series were perfectly (or perhaps imperfectly) correlated".

This is presented as proof that the plan I am supposed to have advocated is unworkable. But it bears no resemblance to anything I have proposed. In the first place, the dependent variable is not a feature of conscious experience, but an aspect of behaviour. Secondly, the example refers to a system in which there are many individuals, whereas I insisted that it should contain only one person at a time, just as Galileo's equation refers to one ball rolling down one inclined plane. To understand how improper this substitution of n systems for one system is, it is only necessary to apply the authors' technique to Galileo's own problem. We consider not a single ball but n balls rolling down n inclined planes differing in slope, all released at different times, t_r . Let s represent the distances of the n balls from their respective starting points at one and the same moment of time, t_0 , and let $t = t_0 - t_r$. Then of course there will be "no means of discriminating between $s = \frac{1}{2}at$ and $s = \frac{1}{2}at^2$ ". The reason is perfectly obvious. We have transformed the parameter a into a variable, while retaining it in the equation as a constant. Returning to the progress-industry example, we may say that the sources of variation in the effect of industry on progress (the parameters of the individual systems) have been spirited out of the way, and in consequence we end up with a system having all the irregularity that could be desired by one who accepts the axiom of indeterminacy. This is no rational argument, but a conjuring trick.

The choice of an equation from Galileo rather than from Newton is instructive from another point of view. We are told that "Galileo claimed to *deduce* the equation from his experiments with rolling balls". Actually it would be more accurate to say it was arrived at by an application of Francis Bacon's principle of *induction*. In this respect it is poles asunder from Newton's theory of gravitation, which was constructed by an act of creative thought. Kepler had suggested that the planetary orbits were likely to be elliptical, and had postulated an attractive force varying inversely as the distance between two bodies. Bouillaud rejected Kepler's force, and argued that, if it existed at all, it should vary inversely as the square of the distance. Newton invented a new hypothesis incorporating both Kepler's force and Bouillaud's inverse square law; and from this he *deduced* the ellipticity of the orbits.

This brings me to the third major misunderstanding. My critics assume that my reference to Newton implied a recommendation to construct a model for psychology in imitation of Newton's model for celestial mechanics. From this they draw the unwarranted conclusion that I accept the 19th century picture of the physical world. In fact my purpose was to recommend the emulation of Newton's method of approaching his problem. I could just as well have chosen (say) Einstein's 'special theory of relativity', and quoted the Lorentz transformation. (If I had done so, would my critics have deduced that I believe in a universe in which all motion is uniform?) Einstein's *approach* was essentially the same as Newton's. He constructed his theory *first* and tested it afterwards. His first tests were not experimental or observational but logical. In fact it was the logical difficulties inherent in Newton's theory that prompted him to undertake his work. His special theory overcame those difficulties and at the same time admitted the invariant velocity of light, as revealed by the Michelson-Morley experiment. But it was valid only in cases where motion is uniform; and in the next ten years he worked out his 'general theory' to take care of acceleration. It was only *after* he was satisfied with the logical integrity of his theory that he calculated the advance in the perihelion of Mercury, and deduced that the deflection of the light of a star as it passed the sun would be twice as much as his opponents were prepared to admit.

Professor Burt and Mr. Gregory are scornful of the hypothetico-deductive method; and their strictures are fully justified if the hypothesis to be tested has not passed the

rigid logical scrutiny that Einstein thought necessary. But when it has passed that test, the confirmation of a deduction has an immeasurably greater weight. Einstein's theory was fully accepted on the basis of the two deductions mentioned above, and particularly the second one: for so closely knit was the logical structure of the theory that any result other than the predicted one would have been sufficient to prove it false.

Now Galileo's equation, quoted above, does not express a theory at all, but merely summarizes certain observations in a convenient form. It is therefore on a par with the majority of the 'laws' proposed by psychologists—the Weber–Fechner law, Thouless' law of phenomenal regression, and a host of others. There is no point in testing such a 'law' for logical consistency, since nothing can be deduced from it beyond the limited range of phenomena to which it refers. It can only be tested for its accuracy as a description of those phenomena. Yet it is precisely a law of this kind that my critics have chosen to support their argument against me. They are seeking to beat me with a pre-Newtonian stick for my plea that psychologists should emulate Newton's approach to science.

Mr. Pilkington is offended at my complaint that psychologists seldom respect the requirement that a theory should be logically consistent, and would like some examples. I gave one in my contribution to this symposium (p. 136). Here is another. The doctrine of corresponding points is still widely accepted: and the Howard–Dolman threshold, expressed as retinal disparity measured in seconds of arc, is commonly assumed to be a meaningful measure. Yet it can easily be shown, by elementary trigonometry, that if two rods, equidistant from a 'cyclopean eye' are viewed with asymmetric convergence, retinal disparity may be as much as 1000 times the subject's Howard–Dolman threshold without affecting the perceived distances of the rods. Our knowledge of the horopter is as old as the doctrine of corresponding points; yet generations of psychologists have accepted the doctrine despite the fact that the horopter makes nonsense of it.

Here is yet another example. Hebb, in his *Organization of Behaviour*, proposes a theory to account for the visual perception of shape, and illustrates it by showing how it works on a triangle. Fixation of one corner, he assumes, establishes links between certain cells in cortical area 17. The cells thus linked form a 'cell assembly'. Other assemblies are formed when the other corners are fixated. Further scanning of the triangle establishes links between the assemblies: so that, when one is aroused, all are aroused, and the triangle is then perceived as a whole. There is no point in designing an experiment to test this hypothesis, because it can be killed at once by logic. The simultaneous arousal of the three assemblies must give rise to the perception, not of a triangle, but of three lines with a common point of intersection, since the three assemblies have a common point of origin in the cortex. The logical difficulties inherent in Newton's theory pale into insignificance in comparison with this. Newton's theory generated results that were pretty close to the observed facts; his error with respect to the perihelion of Mercury, for example, amounts to an angle of only 42" per century. But Hebb's theory generates results that bear no resemblance whatever to the facts observed. Surely we psychologists need to cultivate something like Einstein's passion for logical consistency.

The question of an axiomatic basis is of fundamental importance; but Mr. Pilkington would maintain that we should be free to choose our own axioms. This I should not for a moment deny. But that is not the same as claiming that one set of axioms is as good as another. Einstein could never have written his general theory of relativity, had he accepted the Euclidean axiom that parallel lines never meet. Similarly, I am convinced that nobody will write a satisfactory theory of conscious processes if he adopts such an axiom as that implicit in the definition of personality given (for example) by Pope Pius XII. Personality he defines (and here he expresses a view that is widely held) as "the psycho-somatic unity of man, in so far as it is determined and governed by the soul" (see *Occupational Psychology*, XXXII, p. 218). But no satisfactory theory can possibly emerge if any lingering fragment of the ghost in the machine is thus admitted. Anyone who would claim that mind is not necessarily lawful will never construct a theory of mind. So, when I say

what must be done, *I* am not presuming to dictate, as Mr. Pilkington supposes: it is the logic of scientific discovery that does the dictating.

I have been censured for my use of the word 'exact'; and I admit that some qualification was advisable. Actually the meaning of the word varies according to our purposes. When people in a particular place witness a predicted eclipse on the day announced, the prediction is 'exactly' confirmed as far as they are concerned, though astronomers may find a discrepancy of a few seconds. For both it is exact in the sense that the probable errors are negligible in relation to the preparations made.

It would be absurd to expect a theory of conscious processes to achieve a higher degree of accuracy than astronomy; but it is surely not unreasonable to demand that it be exact in the sense indicated above, that it should, for example, generate perception of a triangle when a triangle is presented, and not of three intersecting lines. It seems that my critics, while straining at the gnat of the exact meaning of 'exact science', can quite happily swallow a whole cavalcade of camels. The argument of my original article was that we are not likely to arrive at a theory that will achieve even this modest level of exactness if we try to derive it, by Baconian induction or Whewellian consilience of inductions, from masses of statistically ordered data. A theory derived in this way will not account for the strikingly different forms of conscious experience that can be aroused in different people by one and the same situation, or even in one and the same person at different times. Is it intrinsically unreasonable to suggest that we are more likely to succeed if we approach the problem in the spirit of Newton and Einstein? Let me quote from de Morgan's essay on Francis Bacon. "What are large collections of facts for? To make theories *from*, says Bacon: to try ready-made theories *by*, says the history of discovery; it's all the same says the idolater: nonsense, say we! If Newton had taken Bacon for his master, not he, but somebody else, would have been Newton" (*A Budget of Paradoxes*, 1954).

HEBREW PSYCHOLOGY AND CURRENT PSYCHOLOGICAL CONCEPTS

To the Editor, British Journal of Statistical Psychology

SIR,—I should like, if I may, to add my support to the protest which Dr. Taylor has registered against the attempts by Professor Burt and his followers to revive the antiquated doctrine of a 'ghost in the machine' and their frank adoption of a psycho-somatic dualism similar to that which (so Dr. Taylor tells us) was officially sanctioned by the late Pope. I cannot, however, follow Dr. Taylor in his wish to eliminate statistical methods altogether from psychology. Indeed it is because, in common with most present day psychologists and physicists, I believe that statistical methods are basic to natural science that I myself should be prepared to argue that the hypothesis of a *psyche* in a *soma* is by its very nature incapable of verification by the methods of natural science. Moreover, I should agree with Professor Burt that the statistical psychologist and any journal of statistical psychology should devote quite as much attention to defining and expounding the hypotheses to be tested as they do to the processes of statistical analysis and verification.

Where I disagree is over the hypothesis that he and his co-workers have recently been putting forward—or rather resuscitating—in their attacks on behaviourism and on the views of Ryle and the logical positivists. This hypothesis seeks to interpret human behaviour by supposing that each individual person is a composite consisting of three main ingredients. (i) There is first of all the material ingredient, the body or *soma*, and in particular the "material brain—the convoluted lump of nervous tissue seen and stimulated by the surgeon in the operating theatre, dissected post mortem by the anatomist, and inspected under microscope by the histologist"¹. (ii) Secondly, "associated" (as he says) "in some way" with this *soma* is an immaterial entity, the *psyche*—which Professor Burt usually translates by 'mind' and which older writers, of course, would have rendered 'soul'. (In Professor Burt's vocabulary, I imagine, the two are synonymous, 'mind' being preferred because it yields convenient derivatives, e.g. the adjective 'mental'.) This immaterial constituent is invoked more particularly because, in his view, we need something to which we can "ascribe a property which material objects do not possess—the relational property denoted by the phrase 'being conscious of'" (and no doubt other properties as well). (iii) Thirdly, he postulates something that he calls a 'psychic field'. This (if I understand the theory rightly) seems to be conceived as a kind of psychic ether—an immaterial medium which in the neighbourhood of a brain or mind is highly active in highly complicated patterns, and also serves as a vehicle for thought-transference; or, if it is no longer conceived as a psychic ether, it takes the place of such an ether, much as the electro-magnetic field of the physicist takes the place of the old-fashioned physical ether². But in point of fact I can see very little reason for choosing the name 'field' rather than any other: little or no attempt has been made to demonstrate field-properties. Indeed, the concept itself seems to be almost identical with what earlier writers (notably those of an idealist school) would have designated 'spirit'.

Whether or not we prefer the more sophisticated modernized names, the hypothesis in all its essentials seems to be much the same as the traditional conception of man as having a 'tripartite nature' consisting of 'body, soul, and spirit'³. The individual's

¹ Burt, this *Journal*, XIII, 1960, p. 76.

² Cf. Burt, this *Journal*, XII, p. 85, XIII, p. 98, and the (Indian) *Journal of Psychological Researches*, IV, 1960, pp. 54 f., 'The Concept of Mind'. Burt ignores the stock arguments against dualism—that it assumes two kinds of substance and space where one will serve, and thus 'bifurcates nature'; that conscious processes and contents are private and so unverifiable; and that the observation of such processes ('introspection') must inevitably falsify what it claims to observe.

³ Professor Burt himself has commented on the 'analogy', as he terms it.

material body disintegrates after death; but he also possesses an immaterial soul which is then assumed to leave the body, surviving it complete with personal identity, memories, etc., and so continuing to live a life of immortality in a 'spiritual world' (cf. 1 Cor. xv).

And this brings me to the specific argument that I wish to add to those adduced by Taylor, Hansel, Hammond, and others, namely, that these dichotomous and trichotomous conceptions of the nature of man are essentially Eastern, and therefore quite incapable of fitting into the ideational framework of Western science. The tripartite theory is specifically a Jewish doctrine, set forth in the Old Testament writings, taken over almost automatically by the Christian writers of the New Testament, and summed up most succinctly in a single verse by St. Paul (1 Thess. v, 23). The doctrine seems at once natural and familiar to English writers because the impressive translation of those scriptures produced under King James and the conceptions it conveys have so coloured English literature and contributed so deeply to the English language. Thus we find psychologists like William James in America and William McDougall in England putting forward much the same view, and obtaining from their readers a very ready acceptance for it, such as they certainly have not obtained from continental readers, simply because their English readers have, perhaps quite thoughtlessly and almost unconsciously, become used to such ways of thinking from their very childhood. Nevertheless, the concepts involved are, every one of them, elements of an alien and an outworn religious dogma.

My point then is that this hypothesis owes its plausibility, not so much to the explicit arguments and evidence urged in its defence, but rather to the implicit preconceptions which already favour it, and that, when we try to formulate that hypothesis more precisely in empirical terms, we find that, like so many oriental notions, it becomes elusive and ambiguous, and seems to have sprung from a cultural background wholly different from that of the modern scientific world, which is Greek rather than Hebraic. The criticism I am here advancing is not at all new. It was put forward more than once by Watson and the earlier behaviourists in their criticisms of the 'animistic psychology' of McDougall and James, when they pointed out that "the conception of a soul or mind—an immaterial and supernatural anomaly—is a concept of religion not of science, of Hebrew theology not of modern physics".

But professional psychologists are by no means the only writers who have felt that there is something incongruous in the intrusion of such nondescript notions into scientific theories. Quite recently Miss Marghanita Laski, "speaking as a member of the educated public", asked "why the scientific psychologist, if he is not content with interpretations in terms of the brain and nervous processes, cannot introduce more appropriate scientific terms for his explanatory ideas, instead of preserving the ancient designations handed down to us by religious writers, with or without a more or less arbitrary twist to their original meanings".

I would therefore suggest that, if Professor Burt wants us to give serious consideration to the traditional conceptions of mind, soul, and spirit, he should borrow a leaf from the modern logical positivist, and begin by subjecting those terms and notions to a *linguistic analysis*, seeking to determine what precisely is the unique contribution (if any) which they contain within themselves, I personally should expect that they would probably prove to be mere popular circumlocutions for particular processes of the brain and nervous system, invented for or by people who knew nothing of such organs or their functions. If, however, they do contain some other irreducible element, let us have a specific name for it, and a specific definition, so that we can then, if we wish, incorporate it in a tentative hypothesis, and subject it to experimental, statistical, and other empirical tests. But, pending some such unexpected evidence, practically every psychologist and every philosopher must remain firmly convinced that a hypothesis which subdivides the individual personality into two or more separable ingredients—a material body and an immaterial soul capable of independent existence—is altogether untenable. We must at all costs preserve the fundamentally synthetic character of human personality and reject any such analysis or subdivision.

WILLIAM G. RACKHAM.

HEBREW PSYCHOLOGY AND CURRENT PSYCHOLOGICAL CONCEPTS A REPLY

By CYRIL BURT

I heartily agree with Mr. Rackham's plea for an analysis of the meanings of such popular terms as mind, soul, and spirit. It is, as he says, surprising that hardly any of the analytic philosophers have attempted such a task. G. E. Moore has offered a brief analysis of 'mind'¹; and Ryle has discussed the popular uses of many kindred mental phrases. But, since the most familiar source from which such terms as mind, soul, and spirit have entered into current thought is, as Mr. Rackham points out, the Bible, the most instructive approach, it would seem, is to examine the Hebrew and Greek originals, and deduce so far as possible, what precisely is the nature of the primitive ideas which those terms were intended to convey. This will incidentally enable us to check the argument that Mr. Rackham himself has brought forward, namely, that because these conceptions are essentially of Jewish origin, they are therefore inappropriate to the scientific outlook of the Western world, which is essentially of Greek origin.

Since in Hebrew more than one word is employed to express what to the English reader seems almost the same idea, and since each word has several different meanings, often changing with the date of the writings, something like a statistical analysis would be extremely helpful². It is too worth reminding the reader that classical Hebrew is very poor in adjectives and abstract nouns, while the line of division between noun and verb is much less rigid than in Greek or English. And the frequent use of quasi-concrete nouns to express ideas which in English would most naturally be expressed by adjectives, abstract nouns or verbs is apt to be highly misleading.

1. In the Hebrew scriptures by far the most important word for us to study is *nephesh* (נֶפֶשׁ) usually translated 'soul' or 'life'. In the Old Testament it occurs 754 times. 'Soul', the commonest English translation, is scarcely capable of conveying its elusive meaning to a modern reader. The root is found in all Semitic languages. Thus in Arabic *nafasun* means 'breath' and *nafsun* is generally rendered 'soul'.

(i) Most modern scholars (Albright, G. R. Driver, Dürr, Widengren, etc.) seem now to hold that the original meaning of the word was 'throat' or 'neck'. This is the meaning of the old Accadian (Assyrian) word *napista* and the Ugaritic *npš*. And it seems to survive in Isaiah 5¹⁴—a poem which preserves other archaisms (cf. Job 41¹³, where the meaning apparently is that the "breath" from the throat of the crocodile "kindleth coals"). (a) In classical Hebrew the basic idea seems to be the principle of life, thought of in the concrete as either 'that which breathes', or 'the breath' and the reek of hot blood (identified with breath). This explains the curious statement: 'the *nephesh* of the flesh is in the blood' (Lev. 17¹¹). Hence too the prohibition against 'eating blood' (Deut. 12²³). The 'breath'

¹ "One of the chief things we mean by saying we have *minds* is, I think, this: that we perform certain acts of consciousness . . ." (*Some Main Problems of Philosophy*, p. 4).

² The accounts of Hebrew psychology given in most histories (e.g., Brett) are both incomplete and inaccurate, and leave an impression of having been produced by writers who were unfamiliar with Semitic cultures and unable even to read Hebrew. In what follows I have used the Massoretic text (taking the Soncino edition where available). Much of the supplementary information has been drawn from the very comprehensive *Hebrew-English Lexicon* by Briggs Driver, and Brown (1906). An excellent summary is contained in the chapter on 'Hebrew Psychology' in Peake's *The People and the Book*. Unfortunately both these and the admirable articles in the *Encyclopaedias* are now somewhat out of date on certain points, and require to be checked or supplemented by a variety of more specific sources (to be found in the references given below as each point arises). It may be added that the (American) Revised Standard Version removes many of the misunderstandings caused by the Authorized Version (cf. Gen. 2⁷, 1 Kings 10²⁴, Isaiah 40⁷, 53¹⁰).

departs at death (Gen. 35¹⁸; this should not be translated 'the soul departing', nor yet 'giving up the ghost' as in Jer. 16⁸). Life may be restored by artificial respiration (1 Kings 17²¹, 2 Kings 4³⁴—note the sneezing). And thus in many cases the word comes to mean simply 'life itself' (e.g., 'a life for a life', Ex. 21²³). (b) Later, as *neshamah* began to be regularly used for mere breath, *nephesh* (with or without the addition of some such word as *chayyah* = 'living') came to denote a 'living being'. Both words are used in the older (Yahvistic) version of the creation story (Gen. 2⁷): in later writers these cruder fancies tend to disappear or to be treated as poetical imagery. The foregoing group of allied meanings accounts for a greater proportion of uses than any other meaning, viz. 282.

(ii) By an easy transition the word comes to be used merely to denote the individual man himself (and in such cases, it is to be noted, his visible appearance or body is far more prominent than the mental aspect). (a) Thus it is regularly employed (144 times) in enumerations ("sixteen souls", Gen. 46¹⁸, etc.; usually in the later writings of the so-called Priestly school—"P"). (b) It appears quite frequently as a paraphrase for the personal pronoun, especially in poetry (70 times). In two-thirds of the cases (53) the pronoun would be reflexive: e.g., "he loved him as his own soul" (1 Sam. 18^{1, 3}). (c) Thus, paradoxically enough, a word which originally meant 'life' may occasionally be used of a dead body (with or without *meth*: e.g., Haggaï 2¹³; there are 12 other instances of this usage, all in Leviticus and Numbers—late writings of the priestly school: e.g., Num. 6⁶).

(iii) There are 249 cases in which the word is used with a specifically psychological meaning. (a) In early classical Hebrew by far the commonest is to denote the seat of the emotions or passions, much as *leb* ('heart') usually denotes the seat of the intellectual functions (e.g., 1 Sam. 2³⁵, where the English rendering is again misleading). In earlier writings the stock phrase is "my *soul* rejoices, grieves, desires, abhors, loveth, hateth", etc. (b) Far less frequently, and then chiefly in later writers, it is used to denote the seat of the appetites, i.e., of hunger (Is. 56¹¹), and of thirst (Ps. 143⁶). (c) Occasionally in late writers it seems to be used for the seat of knowledge (Prov. 19³, but here the translation is questionable: see R.V. margin), or even for 'will', 'character', and the like. But in all these instances the precise meaning and often even the original text is highly doubtful. This covers all the uses of *nephesh* in the canonical writings. It will be seen that not one of them corresponds with the interpretation given to the word 'soul' by Mr. Rackham.

2. The next most important word is *ruach* (רוח), usually translated 'spirit'. It occurs, however, only half as often as *nephesh* (viz., 378 times). And this word also has a number of different meanings merging into each other.

(i) (a) The basic idea, as the kindred words in Arabic and other Semitic languages imply, is a 'strong wind' either in the natural or in a figurative sense. This accounts for nearly a third of the uses. E.g., Hos. 13¹⁵ "the east wind (sirocco)—the *ruach* of Yahweh—shall come from the desert". In Ez. 42¹⁶ the word is used for the east side or 'quarter'; and in Job 7⁷ "my life is *ruach*" means simply "my life is a windy, i.e. and empty, thing". (b) The word may be used of the breath of the mouth or nostrils (Job 4⁹, where the word is in parallelism with *neshamah*; cf. also Job 15³⁰); in such cases, however, it means 'breath', not so much as the mere principle of life (exc. perhaps Gen. 6¹⁷), but rather as a source of energy, particularly as coming from the Deity. Thus, especially in poetic parallelisms, it is often a synonym for 'the word of command' (e.g., Ps. 33⁶). In the vision of the valley of dry bones (Ez. 37) a *ruach* comes from the four winds (*ruchôth*) and "breathes upon these slain that they may live".

(ii) Of the psychological uses one of the earliest and commonest is to designate 'the spirit of God', especially as manifested in deliberate energetic activity—an obvious extension of the foregoing meanings. In this sense, however, it is not found in any later writer (D, Je, etc.), possibly because the notion had become discredited by the way it was exploited by the 'false prophets'. (a) Most frequently the word refers to the spirit of God as inspiring the ecstatic state of primitive prophecy (frequently in Samuel, e.g., 1 Sam. 10^{6, 10},

etc.) and even the madness of Saul (1 Sam. 16¹⁵). (b) Later (e.g., in Chronicles and Deutero-Isaiah) it is used of non-ecstatic states, in which the prophet is impelled to give warning or instruction (Is. 48¹⁶, 61¹¹). (c) More generally, though less frequently, it is used of the spirit of God imparting other exceptional powers—warlike energy, impassioned bravery, the strength of Samson, and even the energy of life (Gen. 1², Is. 31³, Ez. 1¹²). (d) Finally, in late writers, such phrases as “Thy (His) Holy Spirit” are simply periphrases for God himself (Ps. 51¹¹, 139⁷, Is. 63¹⁰, 11).

(iii) In later writings *ruach* is used to denote the ‘spirit of man’, conceived as that which breathes quickly or violently in animation or agitation and imparts life and energy, especially when bestowed by the gift of God. It thus includes (a) vigour, vivacity, courage (1 Kg. 10⁵, Ps. 76¹³); (b) anger (Prov. 16³²); (c) a troubled mood or disposition, or an uncontrollable impulse, especially as coming from some unusual or mysterious source (2 Kg. 19⁷, Is. 37⁷); (d) the spirit of life generally in men or animals: this usage, however, is post-exilic, and the word is then commonly used in parallelism with *neshamah* (e.g., Ps. 104^{27f}).

(iv) The permanent substratum of man’s conscious life—this meaning also appears only in late Hebrew. In such cases the word is usually found in parallelism with *leb*, and often carries a moral implication (e.g., Ps. 31⁵, 32², 51¹⁰). After Ezekiel’s time the personal sense predominates, and the distinction between *nephesh* and *ruach* begins to fade: in Is. 26⁹ (a late passage) the two are used in poetic parallelism; and in Eccles. 3²¹ *ruach* is actually used in this sense of ‘beasts’ as well as of ‘man’, but this is an exceptional usage from the pen of a writer who is quite outside the usual tradition. In all these instances, however, we are not dealing with a mere synonym for *nephesh*; much less with a third constituent in a tripartite constitution. We are confronted with a word which has a very different history. Even in its most general applications something of its higher origin as imparted by the Spirit of God and implying a somewhat exceptional state still seems to cling to it.

TABLE 1. FREQUENCY OF MEANINGS

<i>Nephesh</i> (‘soul’)		<i>Ruach</i> (‘spirit’)	
(i) Breath of life, living being	282	(i) Air in motion	
(ii) An individual:		Wind	117
= pronoun	70	Breath	33
= person	144		—
= body or corpse	19		150
	—	(ii) Spirit of God	94
	223	(iii) Spirit of man	101
(iii) Psychological uses:		(iv) Substratum of consciousness	33
Seat of emotions	158		
Seat of appetites	52		
Miscellaneous	39		
	—		
	249		
	—		
Total	754	Total	378
	—		—

3. *Peripheral organs.* The Jew attributed consciousness both to the body as a whole and to almost every part of the body. Knowing nothing of the nervous system or of the way the peripheral organs are linked to the brain, he located the conscious processes associated with those organs in the organs themselves. Thus for him it is the eye that sees, the ear that hears, the mouth or tongue that speaks. The modern reader is apt to take this as picturesque metaphor; but for the Jew himself it was a literal description of the facts. He analyses his own consciousness, not into psychic functions according to

their mental nature, but into the working of certain specific bodily organs. And the organization of a Hebrew personality is rather like the organization of a Hebrew sentence: its parts are just loosely coordinated, not built up systematically according to a hierarchical scheme, as they are, for example, in Aristotle's analysis of the *psyche*. All this incidentally often enables him to regard his feelings and impulses with some detachment, as though resulting from an independent cause, or, as a modern writer might say, from his 'unconscious self'.

4. Quasi-psychical functions are also assigned to many of the inner organs. Under this heading the most important word is *lēb* or *lēbāb*¹, almost always translated 'heart'. The word occurs 851 times, and is thus even more frequent than *nephesh* ('soul'). The original meaning is doubtful. The Assyrian *labātu* denotes something in 'restless movement'. But the earliest uses in Hebrew and other Semitic tongues suggest rather that the word simply means 'the inner, middle, or central part' of anything, not necessarily a particular bodily organ.

(i) It is thus used, though rarely, of the 'heart' of inanimate *things* (e.g., of the sea, Ex. 15⁸, of the heavens, Deut. 4¹¹).

(ii) In the vast majority of cases the word is used of *human beings*. (a) It may denote the inner life or character generally. Nearly a quarter of the cases would fall under this head. It is thus used in parallel with 'inward parts' (Ps. 22¹⁵, etc.), or in antithesis to 'hands' (Ps. 73²³, etc.), or 'garments' (e.g., Joel 2²³, 'rend your heart and not your garments'). (b) More often, however, it serves as the nearest approach to a Hebrew term for the seat of the *intellectual* functions, and is then often translated 'mind'. This accounts for well over a quarter of all the cases. Note that this usage is the opposite of the English: thus when Hosea (7¹⁴) describes Ephraim as 'heartless' he means, not devoid of pity, but devoid of intelligence: in fact, as he puts it, "like a silly dove". More specifically *lēb* (or *lēbāb*) is the organ of knowledge and sound judgement (e.g., 1 Kg. 10²⁴, of the 'wisdom' in Solomon's 'heart'; cf. Prov. 16^{21, 23}), and of thought and reflection (Sam. 13³, Hag. 2^{15, 18}), and of memory (Deut. 49, Ps. 31¹²). (c) It is also used, but far less frequently, to denote the organ of will (Judges 9³, Exod. 14⁵); and (d) more generally of the moral character; thus we read of a "clean, new, pure, upright, or broken heart". (e) Still more rarely it is used to describe the seat of the emotions, rather like *nephesh* (Ps. 104¹⁵, 105³, etc.).

5. It is odd that the *liver* (*kābhēd* כִּבְד, the 'heavy organ') which in Babylonian psychology plays such an important part and which in later Semitic speech (e.g., Aramaic and Arabic) is treated as the seat of the more violent passions (anger and the like) has hardly any place in Hebrew psychology. In Lam. 2¹¹ we read "my liver is poured out upon the earth" in indignation or sorrow; but in the other instances the text is questionable: the same consonants can easily be pointed differently to read כִּבְדָּ (the construct), i.e. 'glory', כִּבְדָּ 'to make rich', etc.

6. *Kidneys* ('reins', *kelayōth*) are mentioned 31 times as organs with mental functions. They seem to be the seat of those particular feelings which determine preferences, and consequently are apt to be 'tried' by Yahweh (e.g., Ps. 7⁹, Prov. 23¹⁶). 'Reins' and 'heart' are often used in parallelism; and according to Rabbinic teaching "it is the heart that examines and the kidneys that advise".

7. *Bowels* (*me'im*) is used about a dozen times in a psychical sense. They are, so to speak, the organ of pity or compassion (e.g., Is. 16¹¹) or of love (Song of Solomon 5¹⁴), sometimes (it appears) the source of procreation.

¹ There seems to be no great difference in the meaning between the two forms as such. *Leb* is found in J, in early strata of E, and in the earliest poetry, and so tends to bear slightly cruder meanings. *Lebab* is first found in Isaiah, and is regularly used in Chronicles and the Deuteronomic writings.

8. Of the two words translated 'loins' *chalāsîm* (=more precisely 'hips') is used to designate the seat of virility and so of the associated feelings; *môthnāyim* (=more precisely 'waist') is regarded rather as the source of strength generally, alike in man and in animals (e.g., the hippopotamus, Na. 2²). Both words are occasionally used to designate the seat of pain.

9. Even *bones* (*atsāmôth*) are named as though they were a source of feeling. They experience fear, sorrow, impatience, joy. And the word is actually used in parallelism with 'soul' (Ps. 35⁹, 10).

From this it will now be obvious that Hebrew psychology implies no simple tripartite constitution, such as Mr. Rackham and others have assumed. The reason for the assumption is clear. At Alexandria the Greek translators of the Old Testament, in rendering the words *nephesh* and *ruach*, selected the terms *ψυχή* and *πνεῦμα*, which later became associated with the tripartite theory; and the Latin and English translators later on did much the same. The Jew drew no sharp distinction between 'body' and 'mind' (or 'soul'). Indeed, not only in Hebrew is there no word for 'mind' or 'soul' in the modern sense; there is also no distinctive word for 'body'. The commonest substitute is *בָּשָׂר* (*bāsār*) which means 'flesh'; yet even this is apt to take on psychical characteristics, being often used to designate man's mental and moral weakness (Is. 40⁶, Ps. 84²)¹. Hebrew psychology, from the earliest period almost to the last, treats the individual personality as a concrete whole, and regards it as virtually identical with that we should call the individual's body. Personal identity is identity of body: "bone of my bone and flesh of my flesh".

Thus, except for the naïve way in which mental functions are attributed to bodily organs, the Hebrew conception of man represents the very type of psychology which Mr. Rackham advocates. The individual organism *is* the person; the physical body itself is conscious, and is responsible for mental activity, or, as Mr. Rackham would prefer to say, conscious behaviour. There is no 'ghost in the machine', no 'soul' or 'spirit' to survive bodily death. Almost throughout the whole of the Old Testament it is assumed that the conscious personality cannot survive death (e.g., Ps. 5, etc.). In two passages only, both extremely late and written as a result of external influences (chiefly Persian) there is a passing reference to a life after death (Is. 26¹⁹—an apocalyptic insertion of uncertain date—and Dan. 12², written in 165 B.C.). It involves a partial and special resurrection only; and consists in a resuscitation of bodily life, the only kind of life the Hebrew could conceive².

In the New Testament the doctrine of a general resurrection appears already fully developed, and plays an important part in the teaching. It is, however, still a resurrection of the body. For the rest, the conception of human personality is in the main essentially that of the Old Testament with a few minor developments, often in part anticipated in the Apocrypha and Pseudepigrapha which appeared during the intervening period.

Of the New Testament writers³ the author who presents the fullest discussion of the nature of man is Paul of Tarsus—a member of the Pharisaic sect, educated at Jerusalem at

¹ The word *gevīyyâh* usually means a *dead* body or corpse, and in any case is only used about a dozen times in the whole of the O.T.

² Neither the doctrine of Sheol, nor the notion of ghosts—or rather doubles—('ôbh, such as that which was called up by the Witch of Endor) or of phantoms which appear in dreams or chirp and rustle (Is. 8⁹, 29⁴) can be regarded as contradicting what is said above. They merely emphasize the fact that, in a library or corpus of books, written at different periods and in different parts of the country, we ought not to look for any explicit and coherent theory. 'Hebrew psychology' is a set of implicit ideas and assumptions drawn from various independent sources—often indeed from early folk lore—and frequently used by later writers simply to yield a picturesque and vivid expression of the semi-poetic meditations they wish to enforce.

³ For more detailed information see H. Wheeler Robinson, *The Christian View of Man*; W. Morgan, *Religion and Theology of St. Paul*; W. D. Stacey, *The Pauline View of Man*, 1956.

the feet of Rabban Gamaliel (grandson of Hillel). It is to Paul, not to the Old Testament writers, that the tripartite division of the human nature—'body, soul, and spirit' is commonly attributed. This, however, implies that Paul adopted the analytic view of man distinctive of the Greeks (as Mr. Rackham assumes); whereas in point of fact Paul's view was the synthetic view of the Hebrews. There is, however, one striking difference which a simple statistical study quickly reveals. In the Old Testament *nephesh* is used 756 times, *ruach* only 378—a proportion of 2 to 1; in Paul *ψυχή* occurs only 13 times, and *πνεῦμα* 146 times—a proportion of 1 to 11. The change of emphasis is important: Paul, it is plain, is not really interested in the soul.

(i) In the Authorized Version of the New Testament the word 'soul' is invariably a translation of the Greek *ψυχή*, which in turn was the stock LXX rendering of the Hebrew *nephesh*. Paul uses it, not in the Greek sense, but in the simpler Hebrew. For him it denotes, not so much 'the principle of animal life', much less the 'immaterial entity which forms the basis of consciousness', but rather, broadly and vaguely, just 'the state of being alive'. The existence of a 'soul' without a 'body' (or as he preferred to say without 'flesh') would have been inconceivable to Paul.²

(ii) In the Authorized Version the Latin word 'spirit' (less frequently the Teutonic word 'ghost') is used to translate the Greek word *πνεῦμα*, which was the LXX rendering of the Hebrew *ruach*. In the Pauline writings it is used primarily of God, of the Holy Spirit (a synonym for the 'Spirit of God'), and of the 'Spirit of Christ': it denotes the dynamic energy of God in creating, redeeming, and sanctifying man. But the largest and most significant batch of uses concern the various effects of this divine power or influence in the life of believers, creating in them new 'spiritual gifts'. This with Paul develops into a new and highly original conception, far beyond anything met with in earlier Hebrew literature. His use of the word itself, and possibly even the conception, seem to owe something to the Stoic doctrine of *πνεῦμα*. Certainly the Stoic phraseology of Eph. 4⁵ and Col. 1¹⁶ strongly suggest that there at least Paul was interpreting Christ in terms of the *anima mundi*³. But Paul's conception is far more personal. It is plainly misleading to say, as is so often done, that "spirit is the highest element in man", or to suppose that it is the property of every man's soul as such.

(iii) *Noûs*, in the Bible usually translated 'mind', sometimes 'understanding', is a term that is wholly Greek. The metaphysical conception of *νοῦς* as the sovereign element in the universe—a conception which we find in Anaxagoras, the later Stoic philosophers, and Philo Judaeus—is quite foreign to Paul. Nor does he use the word to designate a special capacity or faculty (e.g., the ability to reason, or the basis of intellectual activity, or of conscious activities in whatever form). Rather it denotes the *whole man* as engaged in these various cognitive tasks—knowing, understanding, judging, etc.—not unlike the Hebrew *leb*.

(iv) Mr. Rackham refers to the doctrine of the resurrection as expounded in 1 Cor. 15. But this celebrated chapter must be almost unintelligible to the ordinary English reader,

¹ The other meanings are rarer and doubtful, e.g., 'will' in Col. 3²³ and Phil. 1²⁷; 'person' in Rom. 2⁹ and 13¹.

² Neither Paul nor any other Hebrew could say with Mr. Rackham that "an individual possesses a soul", but only that he "is a living soul". It is curious how the Western mind tends to think in terms of ownership or proprietorship.

³ Cf. W. L. Knox, *Some Hellenistic Elements in Primitive Christianity*. Paul, however, unlike Philo, is not attempting to synthesize the most satisfactory elements in Greek and Jewish theory, but merely, we may imagine, to convey in words intelligible to his Hellenistic readers the views to which he himself had been led by his reflections on his experiences on the road to Damascus and in later life. In writing to different readers, he is quite capable of using different terms or using the same terms in somewhat different senses. However, something approaching Paul's use of the term *πνεῦμα* is to be found in the inter-testamental writers (e.g., Wisdom 7²³, etc., and Enoch 20³, etc.).

who has nothing but the Authorized Version before him. Paul's own distinctive teaching turns on the contrast between a *σῶμα ψυχικόν* and a *σῶμα πνευματικόν*. The Authorized Version translates the former by the phrase 'natural body' which seems wholly inappropriate. From the Greek it is obvious at once that Paul is not putting forward Mr. Rackham's doctrine of the immortality of the soul or psyche. As he implies in the following verse (15⁴⁵) the psyche—bequeathed to us by the 'first Adam'—is merely something 'living' as an animal might live; but pneuma—bestowed by Christ—is life-creating (*ζωοποιῶν*). Paul is writing to inhabitants of Corinth; and he therefore strives to express himself by using terms which they would understand. The ideas he borrows for such purposes are perhaps drawn from popular versions of the Orphic and Platonic doctrines and particularly, it would seem, from the Hermetic writings. The adjectives *ψυχικόν* and *πνευματικόν* have no Hebrew counterpart, and *σῶμα* is a conception that is Greek rather than Jewish. In Paul *ψυχή* and *ψυχικόν* both refer to man's life in so far as he is a mere animal; the 'psychic body' is thus not something made out of psyche or soul, but merely the terrestrial body of flesh, blood, and bone. The 'pneumatic' or 'spiritual' body is not a body made out of 'spirit', as is commonly supposed, but a body corresponding to and expressing the spiritual life which the risen individual will presumably lead. Certain of the Rabbis seem to have held a somewhat similar view: they taught that angelic beings had soul and body from a celestial source; animals had soul and body from the earth; man shared something of both¹.

We may then conclude that nothing in the New Testament writings leads us to modify the view reached above, namely, that the widespread notion which Mr. Rackham has so clearly expressed—the notion that Hebrew psychology adopted a tripartite view of personality and accepted the idea of an immortal and immaterial soul which could leave, and live apart from, the material body—rests on an entirely erroneous impression of the biblical writings².

Where then did the Western world get this dualistic doctrine of a perishable material body linked to an immortal and immaterial soul? The simplest answer, I think, would be to say that it got it from Plato³; at any rate it got it from Greek sources rather than Jewish. But the version of the theory to which Mr. Rackham refers—the version which is, and has

¹ Montefiore and Loewe, *A Rabbinic Anthology*, no. 809. Reitzenstein points to occasional uses of the two adjectives (in the Pauline sense) in the Mystery cults and in the liturgy of Mithra (*Die Hellenistischen Mysterienreligionen*, p. 43; cf. also H. A. Kennedy, *St. Paul and the Mystery Religions*, p. 143).

² Mr. Rackham quotes 1 Thess. 5²³ as "succinctly summing up" the tripartite theory "in a single verse". It is concerned with no such statement. We might just as well suppose that the famous Jewish commandment "to love . . . with all thine heart, with all thy soul, and with all thy might" summed up a tripartite theory. The Pauline verse is simply a prayer for the preservation of the *whole* man in every aspect until the expected Parousia. His meaning is somewhat obscured by the English wording; but is clear from the repeated emphasis of the Greek—*ὅλοτελεῖς* and *ὁλόκληρον*. In any case this epistle is the earliest of Paul's extant writings, written (as is clear) long before he had developed the elaborate theory of 1 Cor. 15. For an authoritative exposition of the passage see H. W. Robinson, *The Christian Doctrine of Man*, p. 108.

³ The dualistic theory, with the two constituents not merely separate, but antagonistic—the theory of an immortal soul pre-existent and imperishable, with a moral and quasi-divine nature, imprisoned in a material body as in a tomb and escaping through redemption ('*σῶμα σῆμα*', was the catchword)—all this was taught in the Orphic Mysteries, and, through Pythagoras and the Pythagorean brotherhood, doubtless greatly influenced the more systematic and sophisticated views developed by Plato. The same source, supplemented by elements from the Platonists and Stoics and from Egyptian, Persian, and other religions, went to produce the Hermetic writings, and what were later developed into the Gnostic religions. These formed the background of popular thought, both educated and uneducated, and at the same time seem to have provided a set of

been, best known—is Neoplatonic rather than Platonic. In Alexandria, round about 200 A.D., there was a remarkable revival and readaptation of Plato's metaphysical teaching; and it was about the same date that the theoretical doctrines, characteristic of later Christianity, began to develop. Origen, the Christian scholar, and Plotinus, the pagan philosopher, were contemporaries who lived during their earlier years in the same city; and, as regards philosophical problems, there is comparatively little difference between the two¹.

Origen makes fun of those who believed that material bodies can be recomposed, and, with hardly any reservation² accepts (as the other members of his school also did) the Greek conception of an immortal and immaterial soul, pre-existent as well as indestructible. However, on these topics the orthodox doctrines of the Christian Church, Protestant as well as Catholic, derive not directly from the Alexandrian school but rather from St. Augustine; and it has been rightly said, "Plotinus was the schoolmaster who brought Augustine to Christ"³. It is in Plotinus that we find the first really clear and explicit statement of the supposed tripartite nature of men. The three constituents are *σῶμα*, *ψυχή*, and *νοῦς*, which are commonly translated Body, Soul, and Spirit. Some writers render *νοῦς* by Intelligence or Reason. But these latter words designate analytic and discursive modes of cognition; whereas *νοῦς*, so far as it is cognitive, is synthetic and intuitive⁴. It is essentially concerned with the apprehension of 'values'—Truth (*ἀλήθεια*

concepts and terms of which many Jewish writers during the inter-testamental period and even Paul and the fourth Evangelist occasionally availed themselves. The sharp antithesis between the visible material things 'below' and the invisible spiritual world 'above', which meets us again and again in these last two writers, is definitely Platonic.

Already in 2 Maccabees (c. 100 B.C.) man is described as made up of two dissimilar parts—body and soul; and here and in kindred passages we now have the Greek word *σῶμα* to designate body in contrast to soul: cf. Wis. 8^{19f}. In 4 Macc. 1³² the author classifies emotions into two types—those belonging to the physical life (*σωματικά*) and those belonging to the mental (*ψυχικά*). Now for the first time in Hebrew literature we find the notion of an independent soul existing before the body (Wisdom 8¹⁸, 15⁸, 4 Macc. 13¹³, 18²³) and a marked emphasis on the moral nature of the soul (Wisdom, 14²², 9³, etc.). The words 'soul' and 'spirit' are now often treated almost as synonyms (we find this in Luke's *Magnificat*—"my soul doth magnify . . . and my spirit hath rejoiced . . ."). Moreover, the word 'spirit' or 'spirits' is also used of supernatural powers. During the Exile the God of the Jews became more and more remotely transcendent, and the conception of angelic beings as intermediary agents assumed increasing importance (we note their frequent appearance in New Testament cosmology, while at times even Christ seems to have been regarded as a kind of super-angel). These spirits are explicitly said to possess no body of flesh (as the 'messengers' of the O.T. apparently did): cf. Luke, 24³⁹. And at times the ordinary human being, having shed his body at death, is conceived as living on in much the same state as these angelic beings (cf. the closing chapters of the Apocalypse).

¹ There was a remarkable development in the theoretical side of Christianity about this period. Down to the time of Marcus Aurelius and later it appeared to be little more than an unorthodox branch of the Jewish religion, secretly cherished by a *tenebrosa et lucifugax natio*; with the rise of the Catechetical school at Alexandria it becomes at once scholarly and philosophical. Origen, indeed, was the most learned man of his generation.

² In order not to contradict St. Paul, he adopts the Stoic doctrine of a final conflagration of the material world, after which the soul (or rather the 'seminal Logos' as a principle of individuation) produces a spiritual body true to type. But to all intents and purposes the resurrection of the body became an entirely otiose dogma, virtually ignored by most members of the Alexandrian school.

³ St. Augustine is constantly quoting Plotinus, often by name: see also his encomium on Plotinus, *Contra Academicos* iii, 18 ("In Plotinus Plato lived again").

⁴ The reason why Plotinus prefers the name *νοῦς* to the name *πνεῦμα* (used by his contemporary Origen to express much the same conception) is that for the Greek philosopher *πνεῦμα* was a

or σοφία), Beauty (κάλλος), and Goodness (ἀγαθότης). These for Plotinus are not mere ideals, but basic constituents of Reality (τὸ ὄν), which for him is inherently spiritual; and the various values or 'goods' form a rational hierarchy depending in varying degrees on the 'First Good'¹.

These psychological concepts therefore are unquestionably of Greek and not of Hebrew origin. They have been transmitted to the modern world, not merely by Christian writers (who from the time of Constantine to Galileo included the majority of the scholars and scientists of their day), but also by non-Christian writers—the pagan authors of Rome, Greece and Alexandria, the Arabian philosophers, and the sceptics of the Renaissance. There is therefore nothing distinctively religious about them, as Mr. Rackham supposes. If we discard or modify them, we shall do so, not because they are "elements of an alien and an outworn religious dogma", but for the same reasons that led to their introduction, viz., the study of observable phenomena interpreted by philosophical speculation.

As regards Mr. Rackham's criticisms of my own views I need only say that on many points he has misunderstood what I endeavoured to convey. But this is hardly the place for a more detailed explanation or defence. I hope to return to the subject in a later issue of the Society's general journal.

technical term used by the Stoic writers, who were materialists and consequently insisted that Spirit was itself material—viz., a purified and tenuous form of air. On the other hand, νοῦς and its adjective νοῦτος conveyed the idea (which for Plotinus was of first importance) that the basis of the 'real' world was, not merely 'good' and 'beautiful' (in the Greek sense, i.e., harmonious and orderly), but *rational*: πνευματικόν would hardly carry these implications. This doctrine, adopted by later Christian philosophers, notably Aquinas, proved to be of supreme importance for the development of science. Thus the opening chapters of Kepler's great work (*Harmonice Mundi*) consist largely of arguments, drawn from the Pythagoreans and Neoplatonists (especially Proclus, whom he constantly quotes) and maintaining at great length that the celestial world, as studied by astronomers, is essentially a rational and harmonious system and therefore capable of explanation in terms of mathematical laws. Kepler (*op. cit.*, v, 3) even borrows Plotinus' elaborate comparison between the 'soul' as cognitive subject and the point at the centre of a circle or sphere (*Enn.* 4, 4, 16)—implying a theory of 'perspectives' rather like that more recently developed by Bertrand Russell (*Analysis of Mind*, p. 70f.). And, when Wordsworth invokes—

Wisdom and Spirit of the Universe,
Thou Soul, that art the eternity of thought,
And givest to forms and images a breath
And everlasting motion . . .

and goes on to discuss how the "human soul is thus built up", the sources of his poetical notions are not the Jewish or the Christian religion, but Plato and Plotinus with their σοφία, νοῦς, πνεῦμα, and ψυχή.

¹ Plotinus, *Enn.* 5. 1. 4f. and 6. 2. 17. A slightly modified and still more attractive formulation is given by Plotinus' chief disciple Proclus (cf. esp. *Theol. Plat.*, i, 1f.). It should be noted that this triadic schematism had become almost as obligatory for a Greek philosopher as it was later on for Hegelian idealism or Marxist dialectics. Three, it was said, was the perfect number: it is the smallest number which has a beginning, a middle, and an end. You start with one thing: the most elementary effort of discrimination subdivides it into two; and these are reconciled in the unity of a higher synthesis, which is thus threefold. These triples are the basic constituents of every classificatory hierarchy as they are of every syllogistic argument. Even Aristotle denied the possibility of a fourth dimension to the universe on the *a priori* ground that "all things are three, and three is everywhere: for, as the Pythagoreans declare, the all itself and all things within it are determined by the number three" (*De Caelo*, i, 1). Inevitably therefore both Plotinus, and the Christian philosophers who were his contemporaries, naturally decided that both man and God must each be a Trinity in Unity.

BOOK REVIEWS

Applications of Information Theory to Psychology. By FRED ATTNEAVE. New York: Henry Holt & Co., 1959. Pp. 120.

It is scarcely ten years since Norbert Wiener developed, in a highly original fashion, Fisher's notion that 'information' could be measured, and suggested the possibility of applying a 'mathematical theory of information' to elucidating the processes of neurology and of human behaviour. The mathematical analysis was elaborated still further by Shannon and Weaver, in relation more especially to the problems of communication. Since then the applications of information theory in psychology have accumulated at an ever-increasing rate. In this country (to mention only a few of the more notable contributions) Oldfield has used it in his studies of memory mechanisms and mental schemata; Hick to throw light on the theory of intelligence testing; Berlyne to investigate the relation between conflict and choice time. Sholl and others have discussed its value and limits as a conceptual scheme for interpreting the work of the cerebral cortex. In America experimental applications have been still more numerous: the method has proved extremely illuminating in dealing with mental phenomena that had previously resisted the traditional methods of experimental attack—the perception of patterns, the influence of meaning on comprehension and on memory, the intelligibility of speech, of printed matter, of visual and audible signals, and of coded messages of every type.

Research students dealing with such problems too often fail to realize the special advantages of this relatively new technique. That no doubt is partly because it has not as yet become a recognized part of the elementary teaching in statistics commonly given to psychological research-students in this country, and partly because the students themselves usually find the printed expositions a little too formidable to follow.

Dr. Attneave, who has already made several original and important contributions of his own to the theory of information, has now written an elementary introduction to the subject with special reference to its applications to psychological topics, and thus supplies a long-felt want. Very naturally, however, the author assumes that his readers will approach their special problems from the standpoint of the psychological student in America rather than from that of the British student. In a paper in this journal and elsewhere I myself ventured to draw attention to the analogies between information theory on the one hand and factor analysis and analysis of variance on the other; and as a teaching device I have found it particularly useful to treat information theory as an extension of the more elementary procedures of the factorist, since that is the approach with which the British student is most familiar.

Consider, for example, the general nature of the data with which information theory has to deal in a typical investigation. It consists of (i) a 'set' of *stimuli*, which have called forth (ii) a 'set' of *responses* from (iii) a 'set' of *persons* on (iv) one or more possible *occasions* out of a 'set' which is or should be carefully specified. For this reason, as is clear from Dr. Attneave's discussion, information theory, like factor analysis, is at bottom largely dependent on the theory of 'sets'. In both cases a typical element in the tabulated data will therefore be n_{ijkl} , where the subscripts indicate that the figure entered in the cell thus designated states the number of times (out of a total of N) person i reacted to test-stimulus j with response k , on the occasion of the l th testing. These four 'sets' and this four-dimensional classification are precisely analogous to those which the factorist envisages in his own investigations. In any one particular experiment no doubt he will simplify his task by eliminating one or more of the four variables: he may, for example, deal with test-data obtained on one occasion only, and treat the various responses as scaled measurements for the several persons. In an analysis of variance and in information

theory the same kind of simplification may be introduced. But the study of the problem in this simplified form should in fact be regarded merely as the first step in the investigation of the more complex issues.

"Information," Dr. Attneave tells us, "is defined as that which would reduce or remove uncertainty." It may therefore be measured by the amount of uncertainty which it succeeds in removing. The term 'uncertainty' suggests that we are thinking of our data as something to be obtained from *future* observations; whereas the term 'information' suggests that we are thinking of the data as something already obtained from *past* observations. But in either case the essential statistical characteristics can be measured in the same way, no matter whether the observations refer to the past or to the future. Time is really irrelevant.

Now uncertainty, as Fisher has pointed out, has many of the characteristics of variance. And information theory, like analysis of variance and factor analysis, deals primarily with a 'set' of possibilities: the data employed and the conclusions reached refer in the first instance always to the entire set as such, and not to any particular individual in the set. Corresponding to the total amount of predictable variance (i.e. the variance in the criterion-variable due to the joint influence of the predictor-variables) we then have the total amount of contingent uncertainty, V say (i.e. the uncertainty in the criterion-variable due to the joint influence of the predictor variables); and, by following the method of maximum likelihood, it can be shown that the quantity

$$2(\log_e 2) n V$$

is distributed approximately as chi-square. Thus analysis of 'information' may be regarded as an analysis of contingency measured by chi-square. Furthermore, we may anticipate that uncertainty can be 'partitioned' into components or 'factors' just as variance and covariance are 'partitioned' in the analysis of variance and in ordinary factorial procedures. In particular it is conceivable that two (or more) variables considered in combination might give more information about some other variable (a 'criterion' let us say) than the same two (or more) variables considered individually. This is plainly analogous to the phenomenon of 'interaction' in the analysis of variance.

The factorist reaches his final specification by a series of dichotomies ('bipolar factors', as I have termed them) after the time-honoured manner of 'Twenty Questions'. The procedure is as old as Porphyry's celebrated 'Tree'; the logician's stock example is Bentham's well-known method of 'successive subdivision' as applied to the classification of botanical or biological data generally (e.g. "organic or inorganic?" "animal or plant?" "seed-bearing or non-seedbearing?" "flowering or non-flowering?" and so on: see *Factors of the Mind*, 1940, pp. 108 and refs.). And the principles of classification—the *fundamenta divisionis*—serve as 'predictors'. Much the same procedure is adopted in analysing information (cf. this *Journal*, IV, pp. 196 f.). Moreover, both in the analysis of variance and in the analysis of uncertainty, deductions become much simpler if the predictors are orthogonal, i.e. exhibit zero correlation; and this is the special condition which factor analysis, at least in its original form, always set out to secure.

Dr. Attneave devotes his longest chapter to the important topic of 'redundance', which has proved to be a most fruitful concept in several different fields of psychological research. Spoken, written, and printed English, as Shannon has shown, generally reveals a 'redundance' (in the technical sense) of about 75 per cent: this of course does not mean that it is to the same extent 'redundant' in the sense of the critic of literary style. The statistical redundancy, as experiments have demonstrated, is a valuable aid to comprehension and to memory: for instance, it appears to be the main, though not perhaps the only, reason why meaningful material can be learnt and recalled so much more easily than nonsense material. And some of Dr. Attneave's own investigations suggest that it may be "one of the essential keys to the correct interpretation of the laws of Gestalt". "Redundancy and organization," he says, "are, roughly speaking, the same." But here

again, I fancy, the argument could be rendered more intelligible to the ordinary student if the basic notion of 'redundancy' were compared with the analogous concept which occurs in the more familiar problems of factor analysis and of multivariate analysis generally, namely, the excess of the variables actually observed over the minimum number of orthogonal factors or components required (i.e. over the rank or number of degrees of freedom).

Since the methods of statistical analysis adopted in information theory are in many ways so similar to those used in other forms of multivariate analysis, the psychological student is naturally led to ask when or why one type of procedure should be chosen rather than the other. This is a question which Dr. Attneave does not discuss in any detail. The choice depends partly on the characteristics of the observational data and partly on the assumptions that the investigator is prepared to assume in his endeavour to verify or refute one or other of the alternative hypotheses which his research has been planned to test. Both analysis of variance and factor analysis as ordinarily used imply some kind of metric; and, if it seems desirable to preserve this metric, then one or other of these modes of analysis should be used. On the other hand, the method of information theory requires no assumptions to be made in regard to metric. It can thus be used to compare analyses of data obtained in two or more investigations where the metric involved differs in the different cases. Shannon's method of measuring 'information', and the theorems deduced from it, rest on certain quite specific assumptions of their own. These are usually fulfilled in telephonic engineering; but they are by no means so often fulfilled in biological and psychological systems. Hence his particular method should be used by psychologists with greater caution than has commonly prevailed. It should be adopted only after a formal demonstration has been undertaken to show that the necessary postulates may reasonably be accepted.

Considering what a slender volume his introductory manual makes, Dr. Attneave's treatment is remarkably comprehensive. The style is as logical as it is lucid. And the book may be unhesitatingly recommended to all research students and instructors who happen to be interested in problems of statistical psychology.

CYRIL BURT.

Reading and the Psychology of Perception. By HUNTER DIACK. Nottingham: Peter Skinner Publishing Limited, 1960. Pp. 155 + xxiii. 21s.

In considering a book like this the readers of a statistical journal will be interested more especially in two points of technique: (i) how far are the author's statistical methods appropriate and valid, and (ii) how far does he try to answer by mere verbal argument questions that can only be solved by a statistical inquiry? Since Dr. Diack's criticisms of previous investigators, and his claims for his own special methods of teaching, have attracted considerable attention both among teachers and in the educational press, it is important to ask whether the data provided by his own investigations are really adequate to support his rather sweeping conclusions.

In one of his chapters Dr. Diack summarizes a statistical study, in which he endeavoured to compare the efficiency of two methods of teaching reading—the 'mixed method', which he regards as 'normal in English schools today', and a 'phonic method' favoured by himself and his collaborator, Mr. J. C. Daniels. Two batches of pupils from four junior schools were selected; one was taught by the mixed method, the other by the phonic method. At the end of a year both were subjected to a test of word recognition. With the former the average proportion of words correctly read was only 43 per cent; with the latter 75 per cent. Dr. Diack says that this demonstrates the complete superiority of the phonic method over the mixed.

Now for such a comparison to throw light on the relative efficiency of the methods used, it is essential that the influence of all other factors be eliminated or at least held constant. Where reading is concerned the most important factors are (i) previous schooling, (ii) the children's intelligence, (iii) the cultural level of the home, and (iv) the enthusiasm of the teacher for the method he is using. (i) Both groups had already had "two years in infant schools using mixed methods", so that the natural inference would be that a mixed method followed by a phonic is best for most children, not that the phonic is wholly superior to the mixed. (ii) "It seemed to us that the factor intelligence was taken care of because about three-quarters of the age group had learned to read under the same school conditions". But this surely provides no guarantee that the two batches were really matched in regard to intelligence. (iii) The pupils selected were those who had made no score in a preliminary word recognition test. This, it is argued, must "largely cut out the influence of the home". But once again this is a very inadequate way of ensuring that the two groups were equally matched in regard to home conditions. (iv) When an outside investigator chooses schools where the teachers are "eager to cooperate", it is always highly likely that those teachers will themselves be keenly interested in the new method that is being proposed. The only safe way to rule out this and all the other irrelevant influences is by combining the method of randomized selection with that of paired matching. Now, in spite of the vast number of investigations on the teaching of reading, I know of only one where this has been employed—the joint study of Burt and Lewis. And finally what is the criterion on which Dr. Diack relies? It is a test of the recognition of unfamiliar words presented in isolation—a test which notoriously favours children taught by phonic methods. It is quite conceivable that, had they been tested for their ability to understand the meaning of complete sentences, the pupils who had been taught by the mixed method would have proved superior to the others.

The fact is, Dr. Diack himself has little patience with quantitative techniques. He places far greater reliance on direct experiments, such as those he here describes in some detail, carried out "with several hundreds of adults". Single words were briefly exposed to demonstrate "the processes by which words are perceived". In the first series the subjects were asked to say *what* they saw without further clue; in the second, they were *told* which word would be exposed; in the third, the word included misprints, and the subjects were asked *how much* they saw.

The following conclusions are drawn. (i) "If the word is presented for a period of the order of one-hundredth of a second it will be seen as a *grey band*". (ii) "There is an exposure-time during which the subject can see no more than *one or two letters*." (iii) "A subject who is dependent solely on the retinal image and his previous knowledge of the word will, as the exposure-time is increased, distinguish *parts of the word*". (iv) When told the word beforehand, "he will 'see' *the word in its entirety*", even though in fact it was not presented in its entirety, but merely "bears a close resemblance to what he is expecting". Dr. Diack regards this as "perhaps the most satisfactory type of experiment . . . There is no question of probability, but of certainty. There is no need for elaborate statistical tables." The experiments "are so simple and straightforward as to be nearly incredible in a field . . . where belief is so highly correlated with the ability on the part of the writer to blind people with science".

As stated some of his conclusions are indeed incredible. To begin with, Dr. Diack, in proving his points, adopts the old pre-experimental notion that, even with short exposures, the duration of the retinal image corresponds with the duration of the stimulus. But, as knowledge of modern cinematography should be sufficient to show, the retinal image resulting from a very brief exposure still lasts an appreciable length of time. Assuming the illumination is adequate and the exposure too brief to permit eye-movements, the actual length of the exposure has little or no effect on what is perceived; and in fact, as Woodworth observes, "the exposure can be as short as we wish". The subjects naturally

require a little training with the apparatus to fixate correctly and avoid blinking at the crucial moment. But under proper experimental conditions, it is commonly found that most adults can correctly report between 6 and 7 letters, no matter how short the exposure. Moreover, recent investigations have shown that the number of letters or the amount of the word actually reported nearly always underestimates what was actually seen: e.g. one observer in McDougall's laboratory, who reported only 5 out of 9 letters shown, remarked: "*All nine were equally clear; and I could have remembered them all, if my reply could have been instantaneous*". And similar observations are reported by other investigators. In short, such experiments merely test how much is recalled, not how much is perceived. If therefore some of Dr. Diack's observers really saw a 'grey band' instead of a row of letters, that can only have been because of some peculiarity in the experimental conditions: either the card was seen from a distance too great for discrimination (as in Wiegand's well-known experiments), or the observers were not adequately trained in the task of fixation, or possibly the illumination was too weak. Unfortunately we are not told any of the technical details—e.g. what kind of tachistoscope was used, whether the "hundreds of adults" were tested individually or in groups, what was the strength of the illumination, or whether the reports were oral or written. Dr. Diack vigorously criticizes the accepted results of earlier laboratory specialists in this field. But, since his own account is never so precise as theirs, few experimentalists, I imagine, will take his claims very seriously¹. May we remind him of Woodworth's warning? "One of the main difficulties in a tachistoscopic research is to make sure of stating all the conditions which another experimenter would need to repeat the experiment."

The most serious error in Dr. Diack's conclusions arises from the implication that the conclusions are virtually the same for all individuals. He speaks continually of '*the child*', as though the mental abilities and processes of all children were the same; and in his closing chapter he insists that "every child who learns to read, unless he is blind or deaf, passes through the same stages"². In point of fact almost every other investigator reports wide individual differences, both quantitative and qualitative, in abilities and in development. The average tachistoscopic span of some children, for instance, may be only 2 to 3 letters, that of others, even of the same age, may be 8 or 9. There are still wider differences in the way different observers attack their task. Meumann in Germany and Burt in this country, in their tachistoscopic work with school children, found every possible type of 'progressive apperception' from the extremely analytic to the wholly synthetic or impressionistic. And plainly conclusions drawn from experiments like those of Dr. Diack which are carried out with adults (whose age, intelligence, and educational level are never stated) cannot be applied without further confirmation to children of school age; nor can results obtained with isolated words be applied as they stand to the processes involved in reading complete sentences on the printed page.

The greater part of Dr. Diack's book is taken up with an uncompromising attack on Gestalt theory and an attempt to substitute a new 'psychology of perception', which he outlines at some length. "For upwards of thirty years" (he tells us in his opening chapter) "writers on the teaching of reading have leaned upon Gestalt: they have continued to do so even after the Gestalt system has crumbled under logical analysis supported by more

¹ Dr. Diack appears rarely to have studied the original articles which he cites and often criticizes. Most of his quotations are at second hand, usually it would seem from Woodworth's *Experimental Psychology*; and he frequently misinterprets the actual work (e.g. in his criticism of Wagner's conclusions, which seems plainly to indicate that he has not read Wagner's own papers).

² He even tells us what the five stages are. But in point of fact, if we may trust the statistical analyses made of representative samples, the type of development he describes is characteristic of only about 27 per cent of the ordinary child population. Even without accepting these and similar percentages, we can safely say it is not characteristic of "*every child*".

precise experimental work than any carried out by the founders of Gestalt theory". It is the Gestalt theory, so he supposes, that has been responsible for the 'Word-whole' or 'Look-and-Say' method which he unsparingly condemns. Since so many educationists have of late accepted the views which Dr. Diack has put forward here and in his earlier writings, it would seem desirable once again to examine the arguments on which his case is based.

First of all, teachers whose memories go well beyond the last thirty years will readily agree that the advocacy of the 'look-and-say' method as *the* sovereign procedure reached its peak long before any one in this country had ever heard of Gestalt; and it was definitely on the decline by the time the well-known works of Koffka and Köhler were published¹. Secondly, it should be noted that Köhler and Koffka, and other members of their school were largely concerned to combat the old-fashioned associationist theories, which were already obsolescent in this country, but still prevailed at that date both in Germany and in America, where their work was chiefly carried out. This accounts for what now seems to us an exaggerated emphasis on the synthetic aspects of perception and an excessive disparagement of the influence of associations previously learnt. In Britain that battle had already been fought and won by Ward, Stout, McDougall, and their various followers; and it was rather to Stout's doctrine of 'schematic apprehension' than to the doctrine of Gestalt that educational psychologists appealed.² The special merit of the Gestalt school was that they devised and carried out a vast number of ingenious experimental researches confirming and elucidating what with Stout and others were mainly introspective observations or armchair arguments. I know of no "more precise experimental work" than the best of their studies. Thirdly, both Dr. Diack's account of the main principles of Gestalt psychology and his claim that it has now "fallen into disrepute" are misleading. Professor Murphy's chapter on the subject is far more accurate and informative. "In general," he says, "Gestalt psychology has been gratefully received and grafted upon existing systems . . . Every nook and cranny of psychology has been invaded with the conception of 'structure'"; and, "though Gestalt theory of course does not offer a final or fundamental solution of all psychological problems", nevertheless its trend and concepts are all "fully in accord with the modern spirit" (*Historical Introduction to Modern Psychology*, 1956, chap. xx).

According to Dr. Diack "it was F. J. Schonell who more than any other single writer lodged Gestalt ideas in the minds of English educationists"; and throughout his chapters

¹ The popularity of the method was largely due to the observation of many enterprising teachers and inspectors (Ballard and Kimmins, for example) that "the intelligent child often uses a 'look-and-say-to-onself' method in spite of the fact that the teacher used some quite different method—the alphabetic or the so-called 'syllabic'": (see B. Dumville, *Fundamentals of Psychology*, 1912). Ruskin, it will be remembered, relates how he "absolutely declined to learn to read by the system of syllabic study" which his teachers tried to impose on him, and instead taught himself by treating "words in their collective aspect", assisted by his "admiration for the sheer look of printed type" (*Praeterita*, I, i). Garlick's *New Manual of Method* (1896), which was one of the books most widely used by teachers in training over sixty years ago, includes a detailed discussion of the look-and-say method and its merits and drawbacks, but eventually recommends a mixed or 'eclectic' method.

² Many of the processes which Dr. Diack seeks to analyse are really processes of apperception. It is his attempt to treat these complex processes as 'simple and straightforward' results of elementary sense perception that accounts for the curious stress he lays on the importance of the mere 'retinal image' and his criticism of the Gestalt psychologists for "paying hardly any attention to the structure of the human eye". (Needless to say, both Koffka and Köhler had made a thorough study of its structure, and would have quickly pointed out mistakes in Dr. Diack's own description.) Moreover it is curious that Dr. Diack confines himself almost entirely to the *visual* perception of words as printed, and makes no attempt to analyse the auditory perception of words as spoken and heard.

there are recurrent attacks on Schonell's book *Backwardness in the Basic Subjects* (1939), which, he tells us, has "long been regarded as a classic". Now, in point of fact one will find far more references to Gestalt in educational books and journals *before* 1939 than afterwards; and, if any single person has "lodged Gestalt ideas in the minds of English educationists", it was surely Schonell's own teacher, Sir Percy Nunn, who claimed that, in accordance with the views of Stout and his followers, he had been teaching the same basic principles "long before the German word *Gestalt* was added to the psychologist's vocabulary". Dr. Diack apparently overlooks the fact that Professor Schonell's book, as its title indicates, is concerned with *retarded* pupils; and Schonell's recommendations seem far more closely in line with those of other educational psychologists who have made special studies of the dull and backward.¹

As to the relative merits of the *pure* phonic and look-and-say methods one of the earliest experiments was that of Professor Valentine, who taught two carefully matched groups to read simple English prose transcribed into the Greek alphabet. Judged by a subsequent test with both familiar and unfamiliar words most normal children did better with the phonic method, but the dull did best with look-and-say. A somewhat similar study by Mr. Dumville confirmed the natural expectation that look-and-say would be best for familiar words and a phonic method for the unfamiliar ('Experiments on Methods of Teaching Reading', *J. Exp. Ped.*, II, 1913, pp. 99-112 and refs.). Many analogous experiments have been carried out since with somewhat better statistical planning, and in the main have fully confirmed these results.

It is plain, however, that Dr. Diack would not agree with such a distinction. In discussing Dr. Morris's report on *Reading in the Primary School* (published by the National Foundation for Educational Research) he gives his own interpretation of the figures obtained. "What the scores actually mean," he writes, "is that those who were of poor intelligence and were given early phonic teaching produced better scores in reading than their intelligence scores would have led one to expect . . . Taking the picture as a whole it would seem that children of poor intelligence who were taught phonics in school progressed unexpectedly well, while many of the best readers were those who had had help with phonics at home." Neither of these statements agrees with the conclusions reached by Dr. Morris; but that, says Dr. Diack, is due to the way an indirect statistical approach was used to "reject results that refused to conform directly to the accepted prejudices". Let us then consider how precisely learning by means of phonic methods was related to the raw scores for reading taken as they stand. The correlations (not cited by Dr. Diack) are as follows: 'phonic instruction with five-year-olds' and 'reading scores', 0.20; 'phonic commencing method' and 'reading scores', 0.09. As the report points out, *neither of these coefficients is statistically significant*. In the second case a larger *hypothetical* correlation (0.47) was obtained after attempting to allow for the *possible effects* of the different levels of intelligence at the different schools. But the correlation which measured these

¹ P. T. Nunn, *Education: Its Data and First Principles* (1930) pp. 216f. According to Burt a *predominantly* look-and-say method is chiefly suited to three main types of children: (i) the innately dull and backward, (ii) those whose visual perception and imagery is stronger than their auditory perception and imagery, (iii) the more experienced readers who are able to take in the sense of sentences and paragraphs quite rapidly without any need to analyse separate words. It may be added that many experienced teachers also find that they can secure greatest progress with children at the initial stages by starting off with look-and-say before proceeding to the more tedious and abstract business of phonic analysis. The obvious defects of the look-and-say method are that it renders little aid in spelling, and it leaves the older and brighter child rather at a loss when he has to deal with new words; but neither of these defects matters greatly with the innately dull, since such children will do little writing, and seldom need to read books with an appreciable number of unfamiliar words (cf. C. Burt, *Mental and Scholastic Tests*, 1921, pp. 283f., and *The Backward Child*, 1937, esp. chap. XIII, 'Perceptual Abilities').

possible effects was itself only -0.19 , and this again was non-significant; thus the hypothetical coefficient deduced by mere computation must itself be non-significant. I suspect it is this last figure that has misled Dr. Diack.

The fact is that the most effective method for the ordinary classroom, as the majority of experienced teachers appear to agree, is the so-called 'mixed method' with free adaptation (so far as the teacher can manage it in the big classes of the ordinary school) to the special needs of each individual according to his abilities and the stage he has reached. In practice this largely means allowing each child to adopt for himself the method that comes most natural to him. At the same time, no matter what method is used, the attitude and enthusiasm of the teacher is always the most important factor. And the vogue of new and exclusive methods—both ancient and modern—and the success which their champions are able to report when they are new, is partly due to the fact that for most of us it is easier to wax enthusiastic over some novel technique that can be labelled and publicized, than over the humdrum principle of using a wide variety of techniques and adjusting them to individual needs: for the latter demands not only enthusiasm, but also experience, patience, and skill.

W. S. HOPKINS

The Gifted Group at Mid-Life. By LEWIS M. TERMAN and MELITA H. ODEN. Stanford, California: Stanford University Press; Oxford: University Press, 1959. Pp. xiii + 187. 36s. net.

This volume, a little slimmer than its predecessors, is the fifth in the series of the reports that have appeared from time to time during the forty years since Professor Terman first started his survey of gifted children. The gifted child was defined as one with an I.Q. of 140 or more; and to discover all who fall into this category the various schools of the five largest Californian cities were systematically scrutinized with the help of the teachers. The teachers' nominees were progressively sifted first with a group test, then with an abridged Stanford-Binet test, and finally with the full Stanford-Binet. This whole procedure yielded 661 boys and girls aged between 7 and 17; and together with supplementary batches, control-groups, and the like, the entire survey covered about 1,500 cases. More than 95 per cent of the group are still participating in the investigation, which will be continued under the provisions made by Professor Terman before he died.

The present volume includes over sixty statistical tables setting forth the present situation. The tables are supplemented by a number of case-histories, chiefly of the less typical instances; and analyses of further autobiographical material are promised in future publications. Wherever possible comparisons are made with figures for the general population or for carefully selected control-groups. Since in the forthcoming *Yearbook of Education* (published jointly in this country and in America) Professor Burt has undertaken a review of the Californian studies, comparing them with similar (but less ambitious) studies in London, New Zealand, and elsewhere, it is unnecessary to embark on a detailed description here.

Professor Terman was always a strong believer in Galton's view that the gifted child is essentially one who is gifted by virtue of his genetic constitution; and the main conclusion brought home by the vast mass of data accumulated in these volumes is that Terman and those psychologists who shared the same beliefs—James, Sully, McDougall, Burt, and their various followers—were correct on all essential points. Recent critics, like Maddox, Stott,

Floud and Fleming in this country and many educational and social psychologists in America would do well to look again at the evidence here available. Indeed the predictive value of the primitive methods used in 1921, when judged by after-careers lasting for nearly forty years, is quite amazing. Still more instructive are the instances where early promise was not fulfilled. Even in the United States official opinion is rapidly changing once again. As an American reviewer states in summing up the present volume: "with today's emphasis on the identification of the gifted, one must be grateful for the vision which has here provided us with detailed information on the factors influencing the achievements of a group typifying the highest one per cent of the school population". J. CONWAY

Child of Our Times. By W. D. WALL. London: National Children's Home, 1959. Pp. 110. 7s. 6d.

Following the precedent set by his predecessors Professor Valentine and Professor Burt, Dr. Wall has expanded the lecture given under the auspices of the National Children's Home into a small book; and the result is an admirable review of the present situation in educational psychology. Since Dr. Wall has spent many years in many different countries, he is able to see the problems that confront this country in a broad perspective; as Director of the National Foundation for Educational Research, he describes them from the standpoint of one who has not only made valuable scientific and statistical investigations of his own, but also planned and supervised those carried out by his colleagues under the National Foundation. The book itself is addressed primarily to those who are actually working in the field of education and have not had the time to acquaint themselves with current trends; but it should also be of value as an up-to-date and well-informed survey even for postgraduate students who wish to have a general background for their more specialized researches. M. HOWARD

A Bibliography of Techniques Used in Survey Research. By W. BELSON and C. R. BELL. The Market Research Society, 1960. Pp. 52. 15s. net.

This is a bibliography of papers and other publications discussing the various technical devices used in sociological surveys, particularly those relating to market research. Merely descriptive papers are excluded. The object has been to bring together those which have "an analytic or evaluative content". Many, we are told, deal with "psychological processes which enter into research operations at the methodological level"; others are reports of experimental work. The list itself is preceded by a short introduction explaining 'how to find particular references' and 'where the articles themselves can be seen'. It is followed by an index classifying the references under various headings and subheadings and by a list of recommended books for further reading. The whole will be of special value to psychological research students interested in various types of survey technique. M. R. JOHNSON

ESP In Relation to Rorschach Test Evaluation: Parapsychological Monographs.

No. 2. By GERTRUDE SCHMEIDLER. Parapsychology Foundation Inc., New York, 1960. Pp. 90.

Dr. Schmeidler's latest research is a continuation of her previous studies of the relation between traits of personality and capacity for extra-sensory perception. Over 1300 persons took part in the experiments, about 150 being tested individually and the rest in class. Extra-sensory perception was tested with randomized sets of cards in much the same way as usual; most of the subjects were then examined with Rorschach slides and with the 'Monroe Inspection Technique' for social adjustment. The data so obtained were subjected to an unusually thorough analysis of variance.

The main conclusions were as follows. Subjects who believe in ESP tend on the average to have higher scores than those who do not; and this tendency is most pronounced for those whose social adjustment is good and for those who show little sign of marked inhibition or marked over-responsiveness. The actual differences were decidedly small; nevertheless, owing to the large numbers involved, the conclusions were fully confirmed at a level of significance lower than $P=0.001$.

W. L. BARTON

Psychological Twin Research: I. A Methodical Study. By TORSTEN HUSÉN.
Stockholm: Almquist & Wiksell, 1959. Pp. 153. Sw. Kr. 18.

Dr. Husén's original aim was to study in a large and representative sample "hereditary and environmental conditions with special reference to intelligence test scores and school marks". This volume is based on psychological, educational, and medical data obtained for nearly a thousand twins discovered among the young men of 20 called up for National Service in Sweden between the years 1948 and 1951. He has devoted special attention to the methodological problems involved in such researches. The book will be of great interest to all our readers; and both methods employed and results obtained raise issues of such importance that we propose to devote a full 'critical notice' to their discussion in the next issue of the *Journal*.

CYRIL BURT

INDEX TO VOLUME XIII

ARTICLES

<i>Author</i>	<i>Title</i>	<i>Page</i>
Baggaley, A. R. and Cattell, R. B.	The Salient Variable Similarity Index for Factor Matching	33
Bock, R. D.	Components of Variance Analysis	151
Burt, C.	Gustav Theodor Fechner: 1860-1960	1
Burt, C.	Logical Positivism and the Concept of Consciousness (Critical Notice)	55
Burt, C.	Factor Analysis of the Wechsler Scale, II	82
Burt, C.	The General Aesthetic Factor, III (Reply)	90
Burt, C.	Statistical Psychology and Physical Laws (Reply)	95
Burt, C.	Experiments on Telepathy in Children: Reply to Mr. Hansel	179
Burt, C.	Hebrew Psychology and Current Psychological Concepts	195
Cattell, R. B. and Baggaley, A. R.	The Salient Variable Similarity Index for Factor Matching	33
Conway, J.	The Intelligence of Twins	104
Dale, H. C. A.	A Study of Subjective Probability	19
Grib, T. F. and Rimoldi, H. J. A.	Pattern Analysis	137
Guttman, L.	The Matrices of Linear Least-Square Image Analysis	109
Hammond, W. R.	Statistical Psychology and Physical Laws	102
Hansel, C. E. M.	Experiments on Telepathy in Children	175
Jackson, M. A.	The Factor Analysis of the Wechsler Scale I	79
Lawley, D. N.	Approximate Methods in Factor Analysis	11
Maxwell, A. E.	Discrepancies in the Variances of Test Results for Normal and Neurotic Children	165
Patterson, C. R.	The General Aesthetic Factor I	87
Pilkington, G. W.	Taylor on Law and Theory in Psychology	47
Rackham, W. G.	Hebrew Psychology and Current Psychological Concepts	193
Rimoldi, H. J. A. and Grib, T. F.	Pattern Analysis	137
Slater, P.	The Analysis of Personal Preferences	119
Stone, M.	An Extension of the Chi-Squared Test for Randomness	31
Taylor, J. C.	Scientific Method: A Reply	189
Williams, E. D.	The General Aesthetic Factor, II	88

BOOK REVIEWS

<i>Author</i>	<i>Title</i>	<i>Page</i>
Attneave, F.	Applications of Information Theory to Psychology	204
Belson, W. and Bell, C. R.	A Bibliography of Techniques used in Survey Research	212
Crombie, A. C.	Turning Points in Physics	107
Diack, H.	Reading and the Psychology of Perception	206
Edwards, A. L.	Experimental Design in Psychological Research	107
Halmos, P. and Iliffe, A.	Readings in General Psychology	108
Husén, T.	Psychological Twin Research	213
McIntosh, D. M.	Educational Guidance and the Pool of Ability	105
Quenouille, H. M.	Rapid Statistical Calculations	106
Schmeidler, G.	ESP in Relation to Rorschach Test Evaluation	213
Terman, L. M. and Oden, M.	The Gifted Group at Mid-Life	211
Wall, W. D.	Child of Our Times	212

THE
BRITISH JOURNAL
OF
STATISTICAL
PSYCHOLOGY

EDITED BY
JOHN WHITFIELD

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN	E. A. PEEL
M. S. BARTLETT	L. S. PENROSE
W. G. EMMETT	J. FRASER ROBERTS
M. G. KENDALL	A. RODGER
D. N. LAWLEY	W. STEPHENSON
E. S. PEARSON	A. STUART
P. E. VERNON	

Printed and Published by
TAYLOR & FRANCIS LTD.

RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20s. per part (U.S.A. \$3.50)

Subscription 30s. 6d. per volume (U.S.A. \$4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30s. net for Members of the Society and 30s. 6d. (post free) for non-members. The subscription price in the U.S.A. is \$4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1: non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, in May and November.

Papers for publication should be sent to JOHN WHITFIELD, University College, Gower Street, London, W.C.1. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 25 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

Books for review should be sent to the Editor, advertisements to the Secretary, British Psychological Society.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology*: subscriptions, 60s. per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology*: for this the subscription, £1 per annum, should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society.

Papers for publication in *The British Journal of Psychology* should be sent to Professor James Drever, Department of Psychology, the University, Edinburgh; those for publication in *The British Journal of Medical Psychology* to Dr. Joseph Sandler, 96 Portland Place, London, W.1; those for publication in *The British Journal of Educational Psychology* to Professor P. E. Vernon, London University Institute of Education, Malet Street, London, W.C.1.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer: that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract: (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the 'Recommendations to Authors' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21-73), or to those given in the paper on 'Setting Mathematics' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost.

In future authors will receive two complete slip-proofs of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

